

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2002, 11:29:25 ; Search time 1907.71 Seconds
(without alignments)
12741.429 Million cell updates/sec

Title: US-09-749-589-1

Perfect score: 2262
Sequence: 1 atgagccagcagccagccgcg.....agacctgacgcctctga 2262

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estom:*
5: em_estol:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: qb_estl:*
11: qb_est2:*
12: qb_hic:*
13: qb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rnd:*
20: em_gss_vtl:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	269.6	11.9	569	10	BE168072 QV3-HT051
2	258.6	11.4	343	11	BG385110 306720 MA
3	189	8.4	837	11	BG743866 602722621
4	188.6	8.3	386	11	BF358998 CM1-ET004
5	179.2	7.9	1246	12	AK018566 Mus muscu
6	170.8	7.6	622	13	AZ833796 2M0116E10
7	158.6	7.0	586	10	BE031984 150773 MA
8	135.4	6.0	857	11	BF578820 602094862
9	132.8	5.9	611	11	BG680597 602629036
10	129.2	5.7	425	10	BE156730 QV0-HT036
11	125.8	5.6	283	10	BB605137 BB605137
12	123.6	5.5	521	11	BG732794 346420 MA

13	122.4	5.4	481	11	BF463964
14	121.8	5.4	538	11	BG727939
15	120.8	5.3	510	10	AA871419
16	119.2	5.3	465	11	BF890309
17	117.8	5.2	524	10	AM656395
18	116	5.1	237	10	BB570880
19	115	5.1	218	11	BF883081
20	111	4.9	595	10	BE156600
21	106	4.7	469	13	AI006678
22	106	4.7	478	13	AQ976378
23	106	4.7	558	10	AI037299
24	105.8	4.7	487	11	BF564736
25	104	4.6	321	10	AM436999
26	103.6	4.6	392	10	AM480438
27	101.6	4.5	385	10	AM142905
28	100	4.4	688	11	BG863674
29	98	4.3	682	11	BF610846
30	97.4	4.3	496	11	BF441578
31	97.4	4.3	497	11	BF441641
32	96.8	4.3	488	10	AA871638
33	93.8	4.1	521	13	AZ066603
34	93.4	4.1	1110	11	BF983692
35	92.8	4.1	568	13	AZ966598
36	89.6	4.0	310	10	AA717241
37	89.4	4.0	716	11	BG718763
38	87.4	3.9	589	10	AM611310
39	87.4	3.9	661	11	B1220929
40	87.4	3.9	784	11	B1330869
41	87.4	3.9	915	11	B1219321
42	87	3.8	556	10	AM603701
43	86.4	3.8	515	11	BG815100
44	85.8	3.8	653	11	B1217880
45	85.6	3.8	799	11	BG966791

ALIGNMENTS

RESULT 1
LOCUS BE168072 569 bp mRNA 21-JUN-2000
DEFINITION QV3-HT0513-150300-117-f10 HT0513 Homo sapiens CDNA, mRNA sequence.
ACCESSION BE168072
VERSION BE168072.1 GI:8630793
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W.J., Zago,M.A., Bordin,S., Costa,F.P., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.O. and Simpson,A.J.

TITLE
JOURNAL
MEDLINE
COMMENT
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
<http://www.ludwig.org.br/scripts/gethtml2.pl?tbl=et2-QV3-HT0513-150300-117-f10&t3=2000-03-15&t4=1>
Seq primer: puc 18 forward

[illegible]

Db	721	GGACCTTGGCTTGGGTTGGCGTCACTCTCTCCCTGCTGCTCGGTTGGGTTCCGGAC	777
RESULT	4		
LOCUS	BF358998/c		
DEFINITION	BF358998	386 bp	mRNA
ACCESSION	CM1-ET0042	130600-264-h04	ET0042 Homo sapiens CDNA, mRNA sequence.
VERSION	BF358998.1	GI:11318070	
KEYWORDS	EST.		
ORGANISM	human.		
SOURCE	human.		
REFERENCE	Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	1 (bases 1 to 386) Das Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. P., Matsukuma, A., Bala, G. S., Simpson, D. H., Brunshtein, A., deOliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.		
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags.		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	97 (7)	3491-3496 (2000)
MEDLINE	20202663		
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?rl=CM1-ET0042-130600-264-h04&f3=2000-06-13&f4=1) Seq primer: puc 18 forward High quality sequence start: 17 High quality sequence stop: 96. Location/Qualifiers 1..386 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="ET0042" /dev_stage="Adult" /note="Organ: Lung_tumor; Vector: puc18; Site:1: Smal; Site:2: Smal; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."		
FEATURES	source		
BASE COUNT	133 a	92 c	92 g
ORIGIN			69 t
Query Match	8.3%	Score 188.6;	DB 11; Length 386;
Best Local Similarity	96.5%;	Pred. No. 6.8e-31;	
Matches 246; Conservative	0;	Mismatches 4;	Indels 5; Gaps 5;
OY	1073	agcagcgtacacgctgatttcgaacccagaaatgatcgctcgtgcgctgcgaacactct	1132
DB	382	agcagcgtacacgctgagtgagtggaacacgaatgatcgctcgtgcgctgcacacacttct	323
OY	1133	ttggtcctctcttaaaatcgaatgcattgctgtgagcttc-tgtcaacttgagctgtg	1191
DB	322	ttggctcctctttttaaattcatgtcatgttggctgtgacgcttctgtgacttggctgtg	263
OY	1192	gatgagcttgagagaaatcccaagttggccacc-tgtgtgtgtctcttggtgtgagt-a	1249
DB	262	gatgagcttgagagaaatcccaagttggccacccttggctgtgtgtgtctcttggtgtgagt-a	203

QY	1250	tcaccatcgtcgtccttggggatctctatcgtatccctccccc-aaagtcctgtg-cttagggac	1307
Db	202	TCACCATCTCTGCTCTGGGGATCTATCTGTATCTCTCCCTTAAGTCTGTCTTAGAGAC	143
QY	1308	ccgtatcgcgtlcaa	1322
Db	142	CCTGATCCCTGCTCA	128
RESULT	5		
AKO18566			
LOCUS			
DEFINITION		Mus musculus adult male colon cDNA, RIKEN full-length enriched library, clone:9030623B18, full insert sequence.	
ACCESSION		AKO18566	
VERSION		AKO18566.1	GI:12858333
KEYWORDS		CAP trapper.	
SOURCE		Mus musculus (strain:C57BL/6J) adult male colon cDNA to mRNA, clone:lib:RIKEN full-length enriched mouse cDNA library clone:9030623B18.	
ORGANISM		Mus musculus	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS		1 (bases 1 to 1246)	
TITLE		Carninci, P. and Hayashizaki, Y.	
JOURNAL		High-efficiency full-length cDNA cloning	
MEDLINE		Methods in enzymology. 303, 19-44 (1999)	
PUBMED		99279253	
REFERENCE		2 (bases 1 to 1246)	
AUTHORS		Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	
TITLE		Normalization and subtraction of cap-trapper-selected cDNAs to	
JOURNAL		prepare full-length cDNA libraries for rapid discovery of new genes	
MEDLINE		Genome research. 10 (10), 1617-1630 (2000)	
PUBMED		20499374	
REFERENCE		3 (bases 1 to 1246)	
AUTHORS		Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasahagi, K., Fujikawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.	
TITLE		RIKEN integrated sequence analysis (RISA) system-384-format	
JOURNAL		sequencing pipeline with 384 multipillarary sequencer	
MEDLINE		Genome research. 10 (11), 1757-1771 (2000)	
PUBMED		20530913	
REFERENCE		4 (bases 1 to 1246)	
AUTHORS		the RIKEN Genome Exploration Research Group Phase II Team and the	
TITLE		FANTOM Consortium.	
JOURNAL		Functional annotation of a full-length mouse cDNA collection	
MEDLINE		Nature 409, 685-690 (2001)	
PUBMED		5 (bases 1 to 1246)	
REFERENCE		5 (bases 1 to 1246)	
AUTHORS		Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arikawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirao, T., Horii, F., Imoto, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sobabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.	
TITLE		Direct Submission	
JOURNAL		Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of	
MEDLINE		Physical and Chemical Research (RIKEN), Laboratory for Genome	
PUBMED		Exploration Research Group, RIKEN Genomic Sciences Center (GSC),	
REFERENCE		RIKEN Yokohama Institute, 1-7-22 Suhei-to-cho, Tsukuba, Iku, Yokohama,	
AUTHORS		kenagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp).	

URL:<http://genome.gsc.riken.go.jp/>, Tel:81-45-503-9222,
Fax:81-45-503-9216)
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAAGGACATTCGAAGACTCTCTTTTCTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot - 10.0 and subtraction to Rot - 185.2. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAAGGACATTCGAAGACTCTCTTTTCTTTTTTTTTTNN 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from lambda FLIC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.

Location/Qualifiers
1..1246
/organism="Mus musculus"
/strain="C57BL/6y"
/db_xref="taxon:10090"
/db_xref="MGD:MGI:1904090"
/db_xref="MGI:MGI:1918793"
/clone="9030623B18"
/sex="male"
/tissue="colon"
/clone_1lb="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
239..1069
/note="putative"
/codon_start=1
/protein_id="BAB31279.1"
/db_xref="GI:12858334"
/translation="MMMPDRCTPRLRSLRRDTLRHSMTSKGVASVLPARIKENTLSIVSGISGLAVYVIOGLAFALNINIPAYCGIAAPFYITIFELGSRHSYGPPEVLSSMAVGVVTRVSDPNASSSELSSSTEDSIEKVAAVSATYLGIQLLGVALQGFVIVLYLSELSIGFTPTAAAIHVILSQLEMLQIPVAYSDFPSIFVLESVFPIOKTNIMADLVTSVIIILVVVEFKFKEINORYRSKLPPVIPIELIMLVIRRLSLII"

BASE COUNT 330 a 298 c 271 g 347 t

ORIGIN

Query Match 7.9% ; Score 179.2; DB 12; Length 1246;
Best Local Similarity 55.4% ; Pred. No. 1e-28;
Matches 393; Conservative 0; Mismatches 308; Indels 9; Gaps 2;

Dn 171 TgTgcTctccTgTgcTcTcccCaagtaCaagaataaaGaactacaTcaTcttcTgaacTgcg 230
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 340 TGTAAGCATCTTGTTGGTTGGCCAGCGCTACAAGATAAAGGAATGGCTTCACATGCATCGTCC 399

Dn 231 tggaactcaaggcggggatccaTccaccaggtcccaCaaggaatgagcatlgtctgcTgcggccaa 290
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 400 TGGCATCAAGCACCGGGCTGTGGCTGTACTCAAGAGATTGTACCATTGGCTGTGGCTGTCAA 459

Dn 291 ccttcctgcaatgaatgTccctctaTcctctcTctcTccccTctgacctgaactcTctcTct 350
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 460 CATCCCTCCACCTTAAGGGGTTGTATGCGCAGGCCCTTTTTCACATTAACCTACTTTTCTT 519

Dn 351 ggaggagtgTtcaacagaatgTgcacagTaactctTgcgcTtlaTcacagaTccTcgtggttaa 410
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 520 GGCGCATCTTAGACACATATCTGTGGGTCATTTCACATTCTTGAGTATGATGAGTGGGAGT 579

Dn 411 catcgtctcagcgtgTgcccaagTcgaaaTcccaagTcttcaCaacatgCcaccacaTga 470
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 580 TGTCGTTACAAGAGTGTCTCGACCCCAA----TGCTCTTCAAGAGTGTCTCA----- 631

Dn 471 gagctatgtgTgaacacagcagTcagTgagagTcgagagTcagTgtcagTgaTcagTcagc 530
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 632 AGTTCACACAAAATGATGCTGTTTCATAGAGCAAAAGTAAVTGGTGGCTGCATCAGTAAAC 690

PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTTCCAGTCACGACG

OY	1365		gggcaccagatataccagggaattaaacatcacgtactgcccttctaacttgcca	1624
Dn	301		AGGCCAAGAAGTCCGGGCGGTGAAGCTTCGCCTCGGCCACCGTGATCTTTGCCA	360
OY	1625		actcagaagtcatc	1640
Dd	361		ATGCTGAATTCTACAG	376
RESULT	10			
LOCUS	BEL56730			
DEFINITION	BEL56730	425 bp	mRNA	EST 21-JUN-2000
ACCESSION	QV0-HT0368	-310300-181-f03	HT0368 Homo sapiens cDNA,	mRNA sequence.
VERSION	BEL56730.1			
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 425) Das Neto,E., Garcia Correia,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.P., Matsushima,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J., and Simpson,A.J.			
AUTHORS	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)			
TITLE	20202663			
JOURNAL	Contact: Simpson A.J.G.			
MEDLINE	Laboratory of Cancer Genetics			
COMMENT	Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel.: +55-11-2704922 Fax: +55-11-2707001 Email: asimpsone@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL http://www.ludwig.org.br/scripts/gethtml2.pl?cl=kt2-Qv0-HT0368-310 300-181-f03&t3=2000-03-31&t4=-1} Seq primer: puc 18 forward High quality sequence stop: 425. Location/Qualifiers 1..425 /organism="Homo sapiens" /db_xref="taxon:9606" /cclone_lib="HT0368" /dev_stage="Adult"note="Organ: head_neck; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription o f tissue mRNA and cDNA amplification were performed under low stringency conditions."			
FEATURES				
SOURCE				
BASE COUNT	77 a	119 c	129 g	100 t
ORIGIN				
Query Match	5.7%	Score 129.2;	DB 10;	Length 425;
Best Local Similarity	59.6%;	Pred. NO. 5.7e-18;		
Matches 218;	Conservative 0;	Mismatches 146;	Indels 0;	Gaps 0;
OY	1275		tctgatcccttcoccttagctgtytctcgtaaggacctgagtcggtcaaatctcaaagaatc	1334
Dn	8		TCTTCATCAACTCCTGCCAACGCGGCTCTTGACCATTCACTCATTTGAAACCTTAAGGCAT	67
OY	1335		cctcaagaactcacaacgcccctactactcgttgtgaagaagcaagctygactgtttgatc	1394

Df		68	GCTGAGGCAGCTCAGCACATCGCTCCCTTGGAAGGGCAATTGCGGGGAGTCTGTTAT	127
Oy	1395	caggagtagtgagcttccctccttccttccttcctaagctcgcccatgtgtygcgaatgg	1454	
Df	128	CttGGGTGGTAGACTTCACGGGCCACCATCTGCTGTAACCTGGACCTTGGCTGGTGGTC	187	
Oy	1455	tgtgcaccttcctcgctcgtgtgctgtctccaagaactagttcgtaaatgataagcaact	1514	
Df	188	GGTCAATCTTCTCCCTCGCTCGTGTGGTGGTGGCGGACAAGAATGCCCACTACTCTGTCT	247	
Oy	1515	ggaccagatcalgatgacaactacatttatgtgaatcccagaacetaataaaggcccaaga	1574	
Df	248	GGGCGAGGTGGCCAGACACAGGATATTTACAGAAATGTGGCAGATCTCAGAGCCCAAGA	307	
Oy	1575	tatccagaggaatatcaatcacagtacagtcgccctcctcraattccaaactagaagat	1634	
Df	308	A GTCCGGGGGGGTGAAGGCTTCCTCGCTCCTCGGCCAACGTTACTTGGCCAATGCTGAGTT	367	
Oy	1635	cttcag 1640		
Df	368	CTACAG 373		
RESULT	11			
Bf605137				
LOCUS	BB605137	283 bp	mRNA	EST 05-DEC-2000
DEFINITION	BB605137 RIKEN full-length enriched, 0 day neonate lung Mus musculus cDNA clone E030008K16 5', mRNA sequence.			
ACCESSION	BB605137			
VERSION	BB605137.1	GI:11556539		
KEYWORDS	EST,			
SOURCE	house mouse,			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 (bases 1 to 283)			
	Alizawa,K., Akahira,S., Akiura,T., Arai,A., Arakawa,T., Carninci,P., Hanagaki,T., Hayasu,N., Hiroaka,T., Hiozane,T., Hodoyama,Y., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J., Koijima,Y., Konno,H., Kusakabe,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi.K., Nomura,K., Numasaki,R., Okazaki,Y., Okido.T., Owa,C., Sakai.C., Seki,K., Sasaki,D., Sato,K., Shibata,K., Shihata.Y., Shinigawa.A., Shiraki,T., Sogabe,Y., Suzuki.H., Tagawa,A., Takahashi,F., Tanaka,T., Toya,T., Watabiki,A., Yamaura,T., Yasunishi,A., Yoshida,K., Yoshihi,A., Muramatsu.M. and Hayashizaki.Y.			
TITLE	RIKEN Mouse ESTs (Alizawa,K. et al. 2000)			
JOURNAL COMMENT	Unpublished (2000) Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel.: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@sc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/ Carninci.P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoke.S., Sasaki.N., Okazaki,Y., Muramatsu.M. and Hayashizaki.Y. Thermosensitization and thermocatalysis of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2). 520-524 (1998) Itoh,M., Kitsumagi,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,K., Ozawa,T., Muramatsu.M., Okazaki.Y. and Hayashizaki,Y. Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5). 463-470 (1999) Carninci,P. and Hayashizaki.Y. High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999) Please visit our web site (http://genome.rtc.riken.go.jp) for further details.			
FEATURES	Location/Qualifiers			
SOURCE	1..283			

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="E030008K16"
/clone_id="RIKEN full-length enriched, 0 day neonate lung"
/tissue_type="lung"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/host="Site_1: Salt; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15', GAGGAGAGACCGCCGCCCACTGCAGATTTCCTTTTTTTTATN 3'}, cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adaptor of sequence 15', GAGGAGAGATTTCCTGCAGATTAAATTAATTAATVCCCCCCCCCCC 3'}. cDNA was cleaved with BamHI and XhoI. Vector: a modified plbluescript KS(+) after bulk excision from Lambda FLG I."

```

Query Match	5.6%	Score 125.8	DB 10	Length 283
Best Local Similarity	80.0%	Pred. No. 2.8e-17		
Matches 148	Conservative 0	Mismatches 37	Indels 0	Gaps 0
QY	1	atgaagccagccagccagcccgctacgltggtagaacaagccgcatctcccttaccctctc	60	
Db	87	ATGAACCCAGCCAGCCAGCCCGCTACGTAATATAGACGTGCGATATCCCTCCCTTC	146	
QY	61	gacgctgagtttgaagaagaagcagcgacatcccaagtggggagaagaactctgcacatgc	120	
Db	147	GATGATATTTTGGAAAAAAGATGAGCTTACCCGGTGGGGAAGAAAGCTTGTAACAT	206	
QY	121	ttcaaatgtttccctcagccaagaatcaaaagctgtggtgttttggactgtcctgtgtctcc	180	
Db	207	TGCGTGTGCTCCTCAGCAAGTTCAAAAGCGTATGTTGGGGGGTGCCTGTGCTTCC	266	
QY	181	tggctc 185		
Db	267	TGGCT 271		

RESULT	12
LOCUS	BG732794
DEFINITION	BG732794 521 bp mRNA EST 11-MAY-2001
ACCESSION	346420 MARC
VERSION	BG732794.1
KEYWORDS	GI:14019077
SOURCE	EST
ORGANISM	pig
REFERENCE	Sus scrofa
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus. 1 (bases 1 to 521)
	Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W., and Keele, J.W.
TITLE	Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
JOURNAL	Unpublished (2000)
COMMENT	Contact: Smith TPL USDA, ARS, US Meat Animal Research Center PO Box 166, Clay Center, NE 68933-0166, USA Tel: 402 762 4366 Fax: 402 762 4390 Email: smtlhemail.marc.usda.gov
	Single pass sequencing. Bases called and alt. trimmed with phred

v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCGCCAGTCACGACG
Plate: 112 row: G column: 5
Seq primer: ATTTAGGTGACACTATG.

FEATURES

source

1. .521

/organism="Sus scrofa"

/db_xref="taxon:9823"

/clone_id="MARC 1P1G"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: PCMV SPOT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from day 11, 13, 15, 20, and 30 embryos."

BASE COUNT 91 a 152 c 149 g 129 t

ORIGIN

Query Match

Best Local Similarity 5.5%; Score 123.6; DB 11; Length 521;

Matches 283; Conservative 0; Mismatches 229; Indels 12; Gaps 1;

377 gtacattgcccgtatcaagcatcctggtgtaacatctgtctgacgtgcccagaagt 436

8 GCACCTTGTGTCATGCTGTGATGAGGAGGAGTGTGACAGAAATCGCTGGCCCCAGATG 67

437 cgaattcccggtgttcaacaatgcccaatgagatagtgtgacacacagccatgg 496

68 AGGACTTCTCAGAGGCTCGAAGCTTAC-----GGTTGATGTGGACGACGACAG 115

497 agagctgagagctgacagctgacagctgacagctgacagctgacagctgacagctg 556

116 ATGCTAGCGGGGTGACAGTGGCCCTGTGACACTAGTGTCTAGTGGCCCTTTTACAGTGG 175

557 gctcggcctcattgagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 616

176 GGGTGGGGCTGCTCAGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 235

617 gctcattgagagcggcgccgctgagatcctgattgtgtgtgtgtgtgtgtgtgtgtgt 676

236 GCTATACACAGCCGCGTCCGTGACAGTCTTCACTCAACCTCAAGTGTGGTGTGGTGG 295

677 tgaccatccctcctacacagagcagagctgacatccttacccttacccttaccctt 736

296 TCCATCTGAGCAGCGCGCTGTGGCCACTGTCCCTCATCTATACAGCACTGGAGGTCTGT 355

737 aaacctcccccac 796

356 GGAATCTGCCACACACTGTACTCAACACCTTGTCACAGTGTGGCAGGTTGGTGGTGG 415

797 tgggtggtgtgtaagagagatcgtcgtacatgacacagatgtgtgtgtgtgtgtgtgt 856

416 TCGGCTGCTGAGCTATTTAACAACAAGCTGCAAGAAACATCTGCTGCGACTGCCCG 475

857 cagagatgattgt 900

476 GGGAGCTACTACGCTCATCGGGGGCACAGGCACTCTCTCTAGGCG 519

RESULT 13

LOCUS BF463964

481 bp mRNA

EST

04-DEC-2000

DEFINITION

U1-M-CG0P-bnr-g-04-0-U1.s1 NIH BMAP_Ret4_S2 Mus musculus cDNA clone

ACCSSION

BF463964

VERSION

BF463964.1

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

AUTHORS

I (bases 1 to 481)

Bonaldio, M.F., Lennon, G., and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: mestrail.nih.gov

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the NotI site

and the oligo-dT track served to identify it as a clone from the

retina tissue cDNA library preparation: M.B. Soares Lab Clone

distribution: Researchers may obtain BMAP cDNA clones from RESEARCH

GENETICS. It should be noted that Bento Soares is generating a

small number of additional specialized non-redundant arrays of BMAP

cDNAs whose availability will be considered under appropriate and

limited collaborative arrangements

Seq primer: M13 Forward

PolyA=Yes.

Location/Qualifiers

1. .481

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="U1-M-CG0P-bnr-g-04-0-U1"

/clone_id="NIH BMAP_Ret4_S2"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pT73d-Pac (Pharmacia) with a modified

polylinker. Site_1: Not I; Site_2: Eco RI; The

NIH BMAP_Ret4_S2 library is a subtracted library,

ultimately derived from mouse retina tissue libraries at

various stages of development. For a detailed description

of the library from which this clone was derived, please

visit our web site at brainest.eng.uiowa.edu.

TAG_LIB=NIH BMAP_Ret4_S2

TAG_TISSUE=adult-retina

TAG_SEQ=GTCCAGCGCGCAC"

BASE COUNT

103 a 124 c 120 g 134 t

ORIGIN

Query Match

Best Local Similarity 5.4%; Score 122.4; DB 11; Length 481;

Matches 210; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

1285 ctccctaaagctgt 1344

59 CTGCCACAGCGCGTCTGCTGCTCTATTTGTAACCTAAAGGATGATGAAGCAG 118

1345 ctccacagccccctactactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1404

119 TTCTGAGACATCTGCTCTTTGGAAGGCAACAGTGGACCTCTATCTGCTGGTGG 178

1405 agct 1464

179 ACCTTTGTGGGCCACAACTCTGCTGAACCTGGAATTTGGCTGGAGTTTCCATATCTTC 238

1465 tccgctcctgt 1524

239 TCTTGTGCTGCTGT 298

1525 atgacactgacattatgtatcctcaagaccataataagagccagatataccagggg 1584

239 CCAAGTACGAGATTTTATAGAGACGTGGCAGATATCTGTGGGCCAAGAGGTGCCGGGT 358


```

/tissue_type="bowel"
/dev_stage="8 weeks"
/lab_host="DH10B"
/notes="Vector: pT713D-Pac (Pharmacia) with a modified
polylinker; Site 1: EcoRI; Site 2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dN) primer [5',
TGTTCGAAATCTGAAGTGGAGGCGCCCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(AATCGGATCCCTTG), digested with Not I and cloned into the
Not I and Eco RI sites of the modified pT713 vector.
Source irradiated bowel harvested 72 hours after
irradiation (1400 Gys). Library constructed by Bob
Barstead."

```

```

BASE COUNT      128 a      130 c      134 g      118 t
ORIGIN

```

```

Query Match      5.3%: Score 120.8; DB 10; Length 510;
Best Local Similarity 58.7%: Pred. No. 4e-16;
Matches 209; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

```

```

OY 1285 ctccctaagtcgtgctagagaccctgacgtgctcaatcctaagaactccctaagcaa 1344
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 25 CTGCCCAAGGCCGCTCTGCTGCTGTCTATTGTGAACCTAAAGGCGATGATGAAGCAG 84
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1345 ctcaacgacccctactactgctgagagaagaagaagctgactgtgcatctggtagtg 1404
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 85 TTCTCAGACATCTGCTCTCTTTGGAGGCAACAGAGTGACCTGCTAATCTGGCTGGTG 144
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1405 agcttcctctctctctctctctctcaagcctgacctgctgagtggtggtgctctc 1464
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 145 ACCTTTGTGGCCACAACTCTGCTGACCTGACATTGGCTGGCAGTTTCCATGATCTTTC 204
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1465 tccgttcctgctgctgctctcctaagaactcagttcgaatagctatgcactgagccaggtc 1524
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 205 TCCTTGCTGCTGCTGCTGCTGCTGAATGCAAGCTGCCCATTTACTCCGCTCTGGGCGCAGGTG 264
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1525 atggaactgacattatgtaaatcccaagaacctataatagggccagagatataccagggg 1584
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 265 CCAGATACGGGCTATTATATAGAGAGCTGCGCAGAAATCTCTGGGGCCCAAGAGAGTCCCGGGT 324
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1585 attaaatcaatcagctactgctccctctctactcttgcacactcagaagatcttcag 1640
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 325 GTGAAGCTTTCGCTCTCTCAGCCAGCGCTGACTTTCGCCAAATGCTGAGCTCTACAG 380
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Search completed: April 25, 2002, 13:42:26
 Job time: 7981 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2002, 12:01:00 ; Search time 98.36 Seconds
(without alignments)
5208.349 Million cell updates/sec

Title: US-09-749-589-1

Perfect score: 2262
Sequence: 1 atgagccagccagccagccgcg.....agaccctgacgcctctgta 2262

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTCUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	326.2	14.4	2882	1 US-08-424-567-1	Sequence 1, Appli
2	326.2	14.4	2882	2 US-08-711-928-1	Sequence 1, Appli
3	326.2	14.4	2882	4 US-09-184-937-1	Sequence 1, Appli
4	120	5.3	588	4 US-09-385-982-129	Sequence 129, App
5	111.4	4.9	623	4 US-09-385-982-513	Sequence 513, App
6	106.2	4.7	475	4 US-09-385-982-146	Sequence 146, App
7	105.2	4.7	595	4 US-09-385-982-477	Sequence 477, App
8	67.4	3.0	583	4 US-09-385-982-491	Sequence 491, App
9	53.8	2.4	7218	1 US-08-237-463-14	Sequence 14, Appli
10	48.4	2.1	816	2 US-07-637-865-1	Sequence 1, Appli
11	48.2	2.1	3489	2 US-08-728-323A-1	Sequence 1, Appli
12	48.2	2.1	32207	2 US-08-770-379-20	Sequence 20, Appli
13	48.2	2.1	32207	4 US-08-757-669A-20	Sequence 20, Appli
14	46.8	2.1	1931	2 US-09-130-114-2	Sequence 2, Appli
15	45.2	2.0	735	4 US-08-998-416-211	Sequence 211, App
16	44.2	2.0	3061	1 US-08-700-576-1	Sequence 1, Appli
17	42.4	1.9	1465	4 US-09-338-671-1	Sequence 1, Appli
18	42	1.9	4403765	4 US-09-103-840A-2	Sequence 2, Appli
19	41.4	1.8	2220	2 US-08-617-801A-1	Sequence 1, Appli
20	40.2	1.8	289	4 US-09-007-005-17	Sequence 17, Appli
21	40.2	1.8	289	4 US-09-244-796-17	Sequence 17, Appli
22	40	1.8	4403765	4 US-09-103-840A-2	Sequence 2, Appli
23	39.8	1.8	1189	1 US-07-781-034-4	Sequence 4, Appli
24	39.8	1.8	1189	5 PCT-US92-08328-4	Sequence 4, Appli
25	39.2	1.7	2712	3 US-09-025-691-4	Sequence 1, Appli
26	39	1.7	1776	1 US-08-484-840-1	Sequence 1, Appli
27	39	1.7	1776	1 US-08-483-094-1	Sequence 1, Appli

28	39	1.7	1776	4 US-09-318-448-13	Sequence 13, Appli
29	38.8	1.7	3300	2 US-08-194-290-6	Sequence 6, Appli
30	38.8	1.7	3300	2 US-08-614-377A-6	Sequence 6, Appli
31	38.8	1.7	3300	4 US-09-142-648B-6	Sequence 6, Appli
32	38.6	1.7	1722	4 US-09-385-028-15	Sequence 15, Appli
33	38.6	1.7	11604	4 US-09-385-028-13	Sequence 13, Appli
34	38.6	1.7	15079	4 US-09-385-028-1	Sequence 1, Appli
35	38	1.7	36519	3 US-08-923-137-2	Sequence 2, Appli
36	37.8	1.7	1491	4 US-09-082-032-9	Sequence 9, Appli
37	37.6	1.7	7295	2 US-08-487-826B-15	Sequence 15, Appli
38	37.4	1.7	20235	1 US-07-642-734C-3	Sequence 3, Appli
39	37.4	1.7	20235	3 US-08-439-009A-3	Sequence 3, Appli
40	37.2	1.6	966	2 US-08-766-738-2	Sequence 2, Appli
41	37.2	1.6	1227	3 US-09-074-912-3	Sequence 3, Appli
42	37.2	1.6	9960	3 US-08-822-586-46	Sequence 46, Appli
43	37	1.6	2156	2 US-08-899-514-1	Sequence 1, Appli
44	37	1.6	3318	4 US-09-593-589-3	Sequence 3, Appli
45	37	1.6	15872	4 US-09-105-537-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-424-567-1
Sequence 1, Application US/08424567
Patent No. 5569755
GENERAL INFORMATION:
APPLICANT: SCHWEINFEST, Clifford W.
TITLE OF INVENTION: Colon Mucosa Gene Having Down-Regulated
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/424,567
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/026,045
FILING DATE: 05-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 40399/181 NIHD
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2882 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 185..2479
US-08-424-567-1
Query Match 14.4% Score 326.2; DB 1; Length 2882;


```

: TELEPHONE: (202)672-5300
: TELEFAX: (202)672-5399
: TELEX: 904136
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2882 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 185..2479
: US-08-711-928-1

```

Query Match	14.4%: Score 32.6; DB 2; Length 2882;
Best Local Similarity	51.7%: Pred. No. 1,4e-71;
Matches	798: Conservative 0; Mismatches 733; Indels 12; Gaps 2.
OY 124	agatgcttcacagccaagaatcaagctgtygtgttggtgctgctgctgctgctgctgctg
Db 314	AGCTGTTCCCAACAAAAGCCAAAGAAATTGGCTCTCTTTGTTCGCCCTACGACTTTGG
OY 184	ctccccaagtacagaattaaagactcaactcctgacctgctgtygagatcagcgcg
Db 374	TTGCCAGCATACCGGCTTAAAGAAAGTTGCTCACTGATGATTTGTTCTGGTATCAGCACA
OY 244	ggatccatcccaaggtcccaagaagcaatgagatctgctgctgctgccaacctctcagtc
Db 434	GGGATTTGGCCCGTACTCAAGATTAGCATTTGGCTGCTGCTGGTGGACATTCGCCCAATC
OY 304	aatggcctctactcctcctctctcccccctctgacctactcttccctgggggtgttcaac
Db 494	TATGGTGTGTATGCACTTCCTTTTCCACCCCAATCTACCTTTCTCTGGCACTTCACGA
OY 384	cagaatgycgaagtgacctgtgcgttcaacatccctgtgtg-----taaca
Db 554	CACATATCCGTCGGTCCGTTTCCGATTTCTGATATGATGTCGGACATGACAGTTTCAGGA
OY 413	tcgtctgcaagctgccccagaatcgaaattccaggtctcaacaatgcc-accaatgag
Db 614	GCAGTTTCAAAAAGCAGTCCCAAGATGCAATGCACTACTTTGGGATTTGGCTTAAACAATCG
OY 472	agctatgtygacacagcagcccatgtagctgtgagagctgcaagtytgcagctacgcctagcc
Db 674	AATATTTCTTACTACTGTGGATGACAGAGAGGTCGAGGGTCGGCGCGGCATATGTCACA
OY 532	tgccctcacccgcatcatccagatggtgctgaggtctcatatgcaattggtcttgtgcccac
Db 724	GTGCTTCTGGAAATCATCCAGTTGGCTTTTGGGATTTCTCGGATTTGGATTTGTAGTGATA
OY 592	tacctctccagatcctcaatccggggtctcatatgacggccggccgtcagatcctgatt
Db 794	TACCTGTCGTGATCCCTCATCATGTCGCTTCACTACTGCTGCTGCTGCTTCATTTTGGTGT
OY 652	tcgggtcccaagtacatctctgagctgacacacccctccctcccaagaagcccaaggtccac
Db 854	TCCCAACTCAAAATTTATTTTTCAGTTGACAGTCCCGTCAACACACTGATCCAGTTTCAATT
OY 712	gtcttaccttcaatltgacaatttgcacaaaacccctccccaacccaacatctgccttcgcatc
Db 914	TTCAAAAGTACTTACTCTGTATTTTCACAAATATGAGAAAGATTAATATTCACAGCTGGTG
OY 772	ttcgctctcatcaagcgtgtgcttctctgtgtgtgtggaagagatcgaatgctcgtcatatg
Db 974	ACAGCTGTGATTTGCTCTTTGGTTGTATCCATTTGTTTAAAGAAATTAATCAAGCGCTTCAAA
OY 832	cacaagatctgcttccccaatcccaacagagatgcatgtgtgtgtgtggaacagctac
Db 1034	GACAAACTTCCAGTCCCATTTCCATTCGATTCGATTCATTTATACCCGTATTTCCACAGAGTGTA
OY 892	tcgggggctgtlaagatgcccacaaagtatcatatgacatcgtyggagaaatccaaagc

```

Dh 1094 TCCATAGGCGTGTACCTTTAAACAGAGTTTAAAGTGGCTGTGGTGGGACATGATTCCT 115
Qy 952 ggggtccccaaccgccggtgcgcctgctgctgcacagtggaagacatgataaggcacagcc 101
Dh 1154 GGATTTTACAGCCCTTATTACACCTGACGTGAGACCTTTCCAAAACACCGTAGAGATTGC 121
Qy 1012 ttccctcctagcaatcgcgagctacgltacatacactgctatggtgcggagaccctggcgcaac 107
Dh 1214 TTTCGGCATCGCAATGAGTTTGACATTTGGACGTGGCCCTTTTCAGTTCCACACCGCTATTTCCCTTC 127
Qy 1072 aagacagcgtacagcagcgtggaattcgaaccagagatgatacgctcctgcgctgcagcaacttc 113
Dh 1274 AAATACGATTATTCACCTTGTGATGGAATCAGAGAGTATATACCTTTGGAGCTGGSTAACTTA 133
Qy 1132 ttgagctcctcctttaaataatcatalytcatttgcctgctgagcttctcgtacactgctgctg 119
Dh 1334 GTCTGTGAGATATTCACAGAGATTGCTGGAGAGTACTGCCCTCTTCACATCAGCATGATTAC 139
Qy 1192 gatgagagctgagagaaaataccacaggtggtccagcctgctgctgctcctgctgagtgatc 125
Dh 1394 GAGAGCACAGAGAGCAAAACACAAATTTGCTGGGCTTATTTGGTGCACATCAGCTGCTGATT 145
Qy 1252 accatgctggtcccgaggagatctatcgtctatccctccctcctaagctcgtgctagagccctg 131
Dh 1454 GTCTGTTCTAGCATTTGGATTTTCTCTTGCGCCCTCTTACAAAAGTCCGCTCGGACGCTTTA 151
Qy 1312 atcgctgtlcaactctcaagaactccctcaagaacatccagaccctactaccctgtgagag 137
Dh 1514 GCATTTGGGAAACCTTAAAGGGAATGCTGATGACAGTTTCTGCAAAATAGCAGATTTGTGGCA 157
Qy 1372 aagagcagaactgagacgtgttcacatctggtgtagtgacatcctcctcctcctcctccagc 143
Dh 1574 AAGGACAAATATGATTTGATTTTAAATTTGGATCATGACATCTTCATCTTCACATTTGCTGGGA 163
Qy 1432 ctgcgccatagtggtgagagtggtggtgcgcctccctccgctcgtgctgctgctccagact 149
Dh 1634 CTCGGGTTTAGGCTCTGGCAGCTAGTGTGGCATTTTAACGTCTAACCATCTGTTTCAGSACC 169
Qy 1492 cagttcgcgaatgctctatgctacgtgcccaggtcactgagacatgacattatgtgaatccc 155
Dh 1694 CAATTTCCAAAATGACAGACGCTGGCTAATATTTGGAAGAACCAACATCTTAAGAAATAA 175
Qy 1552 aagacataatagagccacagatatccagaggatataaatcatcaactgactgctccct 161
Dh 1754 AAAAGATTATTTGATATGATATGACCCAGAAAGACTGAAAATTTTCCATGCTCATCTCCT 181
Qy 1612 ctctactcttgcacactcagagatcttcaagcaaaagtcctg 1654
Dh 1814 ATCTACTTTGCAAAACATTGTTCTTTTAGGGGGAACATTATCG 1856

RESULT 3
US-09-184-937-1
? Sequence 1, Application US/09184937
? Patent No. 6210887
? GENERAL INFORMATION:
? APPLICANT: SCHWEINFEST, Clifford W.
? APPLICANT: PAPAS, Takis S.
? TITLE OF INVENTION: Colon Mucosa Gene Having Down-Regulated
? TITLE OF INVENTION: Expression In Colon Adenomas And Adenocarcinomas
? NUMBER OF SEQUENCES: 8
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Foley & Lardner
? STREET: 3000 K Street, N.W., Suite 500
? CITY: Washington, D.C.
? COUNTRY: USA
? ZIP: 20007-5109
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25

```



```

: GENERAL INFORMATION:
: APPLICANT: ENDEGE, WILSON O., ET AL.
: TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
: TITLE OF INVENTION: PRODUCTS: II
: FILE REFERENCE: CCDNA-260XX
: CURRENT APPLICATION NUMBER: US/09/785,982
: CURRENT FILING DATE: 1999-08-30
: EARLIER APPLICATION NUMBER: 09/328,111
: EARLIER FILING DATE: 1999-06-08
: EARLIER APPLICATION NUMBER: 60/117,393
: EARLIER FILING DATE: 1999-01-27
: EARLIER APPLICATION NUMBER: 60/098,639
: EARLIER FILING DATE: 1998-08-31
: NUMBER OF SEQ ID NOS: 544
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 129
: LENGTH: 588
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(588)
: OTHER INFORMATION: n = A,T,C or G
: US-09-385-982-129

```

```

? CURRENT FILING DATE: 1999-08-30
? EARLIER APPLICATION NUMBER: 09/328,111
? EARLIER FILING DATE: 1999-06-08
? EARLIER APPLICATION NUMBER: 60/117,393
? EARLIER FILING DATE: 1999-01-27
? EARLIER APPLICATION NUMBER: 60/098,639
? EARLIER FILING DATE: 1998-08-31
? NUMBER OF SEQ ID NOS: 544
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 513
? LENGTH: 623
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (1)...(623)
? OTHER INFORMATION: n = A,T,C or G
? OS-09-385-962-513

```

Query Match	5.3%	Score 120	DB 4	Length 588
Best Local Similarity	53.1%	Pred. 9.7e-21		
Matches	255	Conservative	0	Mismatches 225; Indels 0; Gaps 0

QY	1167	tgcgccttcgtacactcgtcgtgtgagatcagcgtgagagaaacccagatggccagcct	1226
Db	6	tgcctctccagatcagcaggttcagagagacagcaggaagcaaacacagatctgcgtcgt	65
QY	1227	gtgtgtgtctctgtgtgtgtgatgataccagatgctgtgtcctgggagctatctatccctc	1286
Db	66	tattgtgtccacacacgtgtcgtgtgtgtgtcgtctctcctcagcatgttatcttccctggcgctct	125
QY	1287	ccctaagatctgtgtctaggagaccctgtatcgtctgtcaactcacaagaactccctaagaact	1346
Db	126	acaaagatccgtctcgtgcagcgtttagcatctgggaacttaaggaaggaatgctgatgagtt	185
QY	1347	caccgaccctactaccctggtgaggaagaagcagcgtgactgtgtcatctgtgtgtgtg	1406
Db	186	tgtcgtgaataagcagatgtgtggtggaagaacaatatgattgttlaattgatcatctgac	245
QY	1407	cttcctccctccctctctctccacgcctccacatgtgtgtgagcagtggtgtgcctctcc	1466
Db	246	cttcatctcaacatgttccctcgtggagctcgtgtgttagccctggagctaggttgcatctca	305
QY	1467	cgctcgtgtcgtgtgtcttcccaagactcagtttcgnaatgtgctatgtcacgtggccaggtcat	1526
Db	306	actgtcaaccacatcgtgtcttagagaccacaaatttccaaatccagacagcgtgtgcaatttgg	365
QY	1527	ggacacgtacatttatgtgaatcccaagaaccataatagggcccgagatccagggagat	1586
Db	366	aagaaaccaacatctctaagaataaaaaagatattatcatgatgatgtatgagccaggaagagat	425
QY	1587	taaaatcatcacgtatcgtctccctctctactcttggccaactcagagatctcaggaanaa	1646
Db	426	gaaattctcagatgtccatctccctactctactcttgcaacacattgtttcttttaggggaa	485

RESULT	5
US-09-385-982-513	
: Sequence 513, Application US/09385982	
: Patent No. 6262334	
: GENERAL INFORMATION:	
: APPLICANT: ENDEGE, WILSON O., ET AL.	
: TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION	
: TITLE OF INVENTION: PRODUCTS: II	
: FILE REFERENCE: CCDNA-260XX	
: CURRENT APPLICATION NUMBER: US/09/385,982	

Query Match 4.9%; Score 111.4; DB 4; Length 623;
Best Local Similarity 52.9%; Pred. No. 1.3e-18;
Matches 258; Conservative 0; Mismatches 22; Indels 1; Gaps 1

QY 1167 tgcgcttcctgctacactcgtgctgagatcgaagctcgtgaggaacaaaccagctggtccagcct 1226
Db 3 tgcctctccagatcagcaggttcagagagacagagaagcaaacacacagatttgcgtgct 62

QY 1227 gtgtgtgtctctgtgtgtgagatgatacccatgtgtgtctccgtggatctactgtatcctct 1286
Db 63 tatgtgtccatcactcgtgtgtatgtatgtcgttctcagcattgtatctctctcgtgcacct 122

QY 1287 cccatgactctgtctatgagagccctgctatgcgtctcaatctcaagaactccctcaagcaact 1346
Db 123 acaaaagtcgctcctcgtgcagcctttagcatcttggaacttaaaaggagatctgtatgtcagtt 182

QY 1347 caccgacccctactactcctgtgtggaagaagcaagcttgagctgttgatctggttagtgag 1406
Db 183 tgcgtgaatagacagatgtgtgtgcgaagaacaatatgttgtttaattgatcatgac 242

QY 1407 ctctctctctctctctctctctcagcctgcacctatgtgtgtgagagtggtgtgcctctcc 1466
Db 243 ctctactctcaccattctctcctcgtgagcctcgtggtcgtgagcgtgagtgtgtgcatctca 302

QY 1467 cgtcctgtgtgtgtgtctccacagccagttctcgaatgtgcatactgtgcacgtgcaggtcat 1526
Db 303 actgctaaccaatcgtgtctcagagacccaattctccaatattgcagacagctgtgcataatttg 362

QY 1527 ggaacactgacatttatgtgaatcccaagactataataggcccgagatataccagggat 1586
Db 363 aagaacccaactctataagaataaaaagatttatgtatgtatgtagccagaagaggt 422

QY 1587 taaatcatcactgatactgtctccctctctacttcttccaactcagagatcttcaggaanaa 1646
Db 423 gaaat-ctcagatgtcatctctcatctactctgnaacactgtnctcttaagcngga 481

QY 1647 ggtcatcg 1654
Db 482 actatcg 489

RESULT 6
US-09-385-982-146
Sequence 146, Application US/09385982
Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: ENDEGE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCDNA-2600X
CURRENT APPLICATION NUMBER: US/09/385,982
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111

```

: EARLIER FILING DATE: 1999-06-08
: EARLIER APPLICATION NUMBER: 60/117,393
: EARLIER FILING DATE: 1999-01-27
: EARLIER APPLICATION NUMBER: 60/098,639
: EARLIER FILING DATE: 1998-08-31
: NUMBER OF SEQ ID NOS: 544
: SOFTWARE: FASTSEQ for Windows Version 3.0.
: SEQ ID NO 146
: LENGTH: 475
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(475)
: OTHER INFORMATION: n - A,T,C or G
US-09-365-962-146

```

Query Match	4.7%	Score 106.2	DB 4	Length 475
Best Local Similarity	56.5%	Pred. No. 2.4e-17		
Matches 262; Conservative	0	Mismatches 189	Indels 13	Gaps 3

```

RESULT 7
US-09-385-982-477
: Sequence 477, Application US/09385982
: Patent No. 6262334
:
GENERAL INFORMATION:
: APPLICANT: ENDEGE, WILSON O., ET AL.
: TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
: FILE REFERENCE: CCDNA-260XX
: CURRENT APPLICATION NUMBER: US/09/385,982
: CURRENT FILING DATE: 1999-08-30
: EARLIER APPLICATION NUMBER: 09/328,111
: EARLIER FILING DATE: 1999-06-08
: EARLIER APPLICATION NUMBER: 60/117,393
: EARLIER FILING DATE: 1999-01-27
: EARLIER APPLICATION NUMBER: 60/098,639
: EARLIER FILING DATE: 1998-08-31
: NUMBER OF SEQ ID NOS: 544

```

```

: SOFTWARE: FastSeq for Windows Version 3.0.
:
: SEQ ID NO 477
:
: LENGTH: 595
:
: TYPE: DNA
:
: ORGANISM: Homo sapiens
:
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(595)
:
: OTHER INFORMATION: n = A,T,C or G
US-09-365-962-477

```

Query Match	4.7%	Score 105.2	DB 4	Length 595
Best Local Similarity	56.1%	Pred. No. 4.5e-17		
Matches 261	Conservative 0	Mismatches 191	Indels 13	Gaps 3

```

? RESULT      8
? US-09-385-982-491/c
? Sequence 491, Application US/09385982
? Patent No. 6262334
?
? GENERAL INFORMATION:
?
? APPLICANT: ENEBE, WILSON O., ET AL.
? TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
? TITLE OF INVENTION: PRODUCTS: 11
? FILE REFERENCE: CODNA-260XX
?
? CURRENT APPLICATION NUMBER: US/09/385,982
?
? CURRENT FILING DATE: 1999-08-30
?
? EARLIER APPLICATION NUMBER: 09/328,111
?
? EARLIER FILING DATE: 1999-06-08
?
? EARLIER APPLICATION NUMBER: 60/117,393
?
? EARLIER FILING DATE: 1999-01-27
?
? EARLIER APPLICATION NUMBER: 60/096,639
?
? EARLIER FILING DATE: 1998-08-31
?
? NUMBER OF SEQ ID NOS: 544
?
? SOFTWARE: FastSeq for Windows Version 3.0
?
? SEQ ID NO 491
?
? LENGTH: 583
?
? TYPE: DNA
?
? ORGANISM: Homo sapiens
?
? FEATURE:

```


ATTORNEY/AGENT INFORMATION:
NAME: Skerpon, Joseph M.
REGISTRATION NUMBER: 29, 864
REFERENCE/DOCKET NUMBER: 0118, 033382
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-296-5500
TELEFAX: 202-296-7830
INFORMATION FOR SEQ ID NO.: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 816 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Flavobacterium
STRAIN: Flavobacterium sp. No. 5942427 141-8
FEATURE:
NAME/KEY: CDS
LOCATION: 1..813
OTHER INFORMATION:
US-07-637-865-1

Query Match 2.1%; Score 48.4; DB 2; Length 816;
Best Local Similarity 53.1%; Pred. No. 0.0061;
Matches 103; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

OY 1799 acggaccaggtgtcctatataccttcagccctgacagctccacactgcccagatg 1798
DB 425 AGGGCGGCTCCCGCGCATCATACCATCGCTCCGTCATTCCTTCATGCGCGG 484
OY 1799 agccacacagctcgcgtcagagcccccagagccagtgacatgctgscagctccac 1858
DB 485 AGGGCGGCGGCTTCATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 544
OY 1899 ccttcgtacactccacacccctacactcgtgacatgagtgagtgacgttcgtgacttga 1918
DB 545 TCGACCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 604
OY 1919 tgggcatcaaggcc 1932
DB 605 CCGGCACACACAC 618

RESULT 11

US-08-728-323A-1/c
Sequence 1, Application US/08728323A
Patent No. 5948676
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
TITLE OF INVENTION: Immediate Early Protein From Kaposi's
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,323A

FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28, 678
REFERENCE/DOCKET NUMBER: 0575/52268/JPM/MS/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO.: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3489 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3489
US-08-728-323A-1

Query Match 2.1%; Score 48.2; DB 2; Length 3489;
Best Local Similarity 46.3%; Pred. No. 0.012;
Matches 158; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

OY 534 cctcacgcccacatccagatggtcgtggtcctacagcttggtggtggtccatcta 593
DB 1288 CCGGCT 1229
OY 554 ccttcagagctcctcctcaggggtcctatgaagcgcgagcctgacatcttcc 653
DB 1228 CCGTCATCTGTCATCTGTCATCTGTCATCTGTCATCTGTCATCTGTCATCTGTC 1169
OY 654 ggtgtcagaatccttcctgagctgacatcccccacacagccaggggtccatcgt 713
DB 1168 GGTCT 1109
OY 714 cttacatcattgacattgcaaaacctccccacacacacatcgtcgtcattct 773
DB 1108 CCGTCATCT 1049
OY 774 cgtctcacaagcgtgctcctcgtggtggtggtggtggtggtggtggtggtggtg 833
DB 1048 GGTCT 989
OY 834 caagattcgtcccccac 874
DB 988 CCGTCATCTGTCATCTGTCATCTGTCATCTGTCATCTGTCATCTGTCATCTGTC 948

RESULT 12

US-08-770-379-20
Sequence 20, Application US/08770379
Patent No. 5849564
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using sw model

```
Run on: April 25, 2002, 13:01:58 ; Search time 214.73 seconds
        (without alignments)
        9031.201 Million cell updates/second
```

Title: US-09-749-589-1

Sequence: 1 atgagccagccagcccg.....agaccctgaccgcccctgtga 2262

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

N_Geneseq_1101.*

1:	/SIDS1/gcgcata/geneseq/geneseqn/NA1980.DAT.*
2:	/SIDS1/gcgcata/geneseq/geneseqn/NA1981.DAT.*
3:	/SIDS1/gcgcata/geneseq/geneseqn/NA1982.DAT.*
4:	/SIDS1/gcgcata/geneseq/geneseqn/NA1983.DAT.*
5:	/SIDS1/gcgcata/geneseq/geneseqn/NA1984.DAT.*
6:	/SIDS1/gcgcata/geneseq/geneseqn/NA1985.DAT.*
7:	/SIDS1/gcgcata/geneseq/geneseqn/NA1986.DAT.*
8:	/SIDS1/gcgcata/geneseq/geneseqn/NA1987.DAT.*
9:	/SIDS1/gcgcata/geneseq/geneseqn/NA1988.DAT.*
10:	/SIDS1/gcgcata/geneseq/geneseqn/NA1989.DAT.*
11:	/SIDS1/gcgcata/geneseq/geneseqn/NA1990.DAT.*
12:	/SIDS1/gcgcata/geneseq/geneseqn/NA1991.DAT.*
13:	/SIDS1/gcgcata/geneseq/geneseqn/NA1992.DAT.*
14:	/SIDS1/gcgcata/geneseq/geneseqn/NA1993.DAT.*
15:	/SIDS1/gcgcata/geneseq/geneseqn/NA1994.DAT.*
16:	/SIDS1/gcgcata/geneseq/geneseqn/NA1995.DAT.*
17:	/SIDS1/gcgcata/geneseq/geneseqn/NA1996.DAT.*
18:	/SIDS1/gcgcata/geneseq/geneseqn/NA1997.DAT.*
19:	/SIDS1/gcgcata/geneseq/geneseqn/NA1998.DAT.*
20:	/SIDS1/gcgcata/geneseq/geneseqn/NA1999.DAT.*
21:	/SIDS1/gcgcata/geneseq/geneseqn/NA2000.DAT.*
22:	/SIDS1/gcgcata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	384	17.0	2580	21 AAD00611	Human membrane tra
2	355.2	15.7	807	22 AA162799	Human CDNA SEQ ID
3	337.6	14.9	443	21 AAC03684	Human CDNA SEQ ID
4	326.2	14.4	2882	21 AA071399	Human secreted pro
5	242.2	10.7	2792	20 AAX78074	Down-regulated in
6	196.6	8.7	1465	22 AA111766	Rat DRDST CDNA. R
7	196.6	8.7	1465	22 AA133076	Probe #1699 for ge
8	196.6	8.7	1465	22 AA101695	Probe #1762 used t
9	192.2	8.5	1521	22 AA120978	Probe #1686 used t
10	192.2	8.5	1521	22 AA146225	Probe #10911 for g
11	192.2	8.5	1521	22 AA106691	Probe #14811 used t
					Probe #6682 used t

12	122.6	5.4	1466	22	AA159022	Human polynucleotide
13	1227	5.4	1227	22	AA160808	Human polynucleotide
14	120	5.3	588	21	AAAI6124	Human colon cancer
15	111.4	4.9	623	21	AAAI6508	Human colon cancer
16	106.2	4.7	475	21	AAAI6141	Human colon cancer
17	105.2	4.7	595	21	AAAI6472	Human colon cancer
18	94.8	4.2	2279	21	AAZ50481	Corn sulphate perm
19	85.8	3.8	1981	21	AAZ50482	Corn sulphate perm
20	80.8	3.6	510	21	AACT4529	Human OREF ORF84 p
21	75.4	3.3	482	22	AAI12132	Probe #2065 for ge
22	75.4	3.3	482	22	AAI13476	Probe #2162 used t
23	73.8	3.3	601	22	AAH08573	Human cDNA clone (
24	73.8	3.3	2403	22	AAH16739	Human cDNA sequenc
25	71.4	3.2	446	22	AAI16178	Probe #6111 for ge
26	71.4	3.2	446	22	AAI13525	Probe #7211 used t
27	68	3.0	165	22	AAI25272	Probe #15205 for g
28	68	3.0	165	22	AAI51390	Probe #20076 used
29	67.4	3.0	583	21	AAAI6486	Human colon cancer
30	66.6	2.9	739	22	AAH33376	Human colon cancer
31	64	2.8	2006	21	AAAC4229	Arabidopsis thalia
32	62.6	2.8	573	22	AAF93562	CDNA encoding SRT
33	61.8	2.7	2241	21	AAZ28466	P. chrysogenum sut
34	58.8	2.6	2449	21	AAZ50491	Soybean sulphate p
35	58	2.6	2022	21	AAZ50491	Wheat sulphate per
36	55	2.4	559	22	AAH96372	CDNA encoding SRT
37	53.8	2.4	345	22	AAH96877	Human EST-derived
38	52.6	2.3	2529	21	AAZ28467	P. chrysogenum sut
39	52.4	2.3	2745	22	AAZ09568	Human transporter
40	52	2.3	2311	22	AAZ50490	Wheat sulphate per
41	51.4	2.3	361	22	AAF65718	Novel human polynu
42	50.8	2.2	696	12	AAO11817	N-acetylmannosamin
43	50.6	2.2	2364	21	AAZ29462	Genomic P. chryso
44	49.8	2.2	445	22	AAH96268	Drosophila EST-der
45	49.8	2.2	1287	21	AACT6396	Human OREF ORF1951

ALIGNMENTS

RESULT

ID	standard; cDNA; 2580 BP.
AAD00611	

AC AAD00611;

DT 29-AUG-2000 (first entry)

DE Human membrane transport protein, MTRP-12 cDNA.

KM Human: membrane transport protein; MRP-12; antiinflammatory; cytostatic;
KM antithyroid; immunosuppressive; thyromimetic; antidiabetic; nootropic;
KM antidiarrheal; neuroprotective; antidepressant; nephrotropic; virucide;
KM antihelminthic; protozoacide; antibacterial; neuroleptic; antiout;
KM diagnosis: prevention; treatment; membrane transport disorder; epilepsy;
KM Alzheimer's disease; diabetes; Parkinson's disease; neurological disorder;
KM inflammatory disorder; AIDS; Addison's disease; atherosclerosis; gout;
KM Graves disease; Hashimoto's thyroiditis; microbial infection; cancer;
KM cell proliferative disorder; ss.

OS Homo sapiens.

FH	Key	Location/Qualifiers
88		88 88 88

PN W0200026245-A2

PD 11-MAY-2000.

PF 04-NOV-1999; 99WO-US26048.

XX

04-NOV-1998; 98US-0172255.
PR 24-NOV-1998; 98US-0172252.
PR 22-DEC-1998; 98US-0172214.
PR 26-FEB-1999; 99US-0121896.
XX
PA (INCYTE) INCYTE PHARM INC.
XX
PI Hillman JL, Yue H, Tang YT, Lal P, Corley NC, Guegler KJ.
PI Baughn MR, Azimzai Y, Lu DAM.
PI
DR WPI; 2000-365576/31.
DR P-PSDB; AAY71067.
XX
XX Novel human membrane transport proteins useful for diagnosis,
PT prevention and treatment of membrane transport disorders,
PT immune/inflammatory disorders and cell proliferative disorders
PT including cancer -
XX
XX Claim 9; Page 119-120; 136pp; English.
XX

CC The present sequence is a cDNA encoding membran transport protein.
CC MRP-12 from Incyte clone 1733477 isolated from human BRST1708 cDNA
CC library. MRP-12 shows homology to mouse transporter (GI 5353730)
CC and human transporter (GI 291964), and is expressed in nervous,
CC gastrointestinal and reproductive tissues.
CC The present sequence is useful in diagnosis, prevention and treatment
CC of disorders related with increased or decreased expression of MRP
CC such as familial goitre, Menkes disease, diabetes, Parkinson's disease,
CC neurological disorders such as Alzheimer's disease, depression, epilepsy
CC schizophrenia, immune/inflammatory disorders such as AIDS, Addison's
CC disease, allergies, atherosclerosis, Graves disease, gout, Hashimoto's
CC thyroiditis, viral, bacterial, fungal, parasitic, protozoal or
CC helminthic infections and cell proliferative disorders such as cancer.
CC Fragments of MRP polynucleotides are useful as targets in microarrays.
CC MRP DNA is also useful for generating hybridisation probes useful in
CC mapping genomic sequences and detecting differences in sequences among
CC normal, carrier and affected individuals. It is also used for
CC screening libraries of compounds in drug screening techniques.

[illegible][illegible]

QY	547	atccagatagggctcgggacttcaatgcagattggcttgggtgcacatctactctccgaatcc	606
Db	611	ttccacgctgaggccgggactgcatccacttcggtctgggtccactatctgcataagaactc	670
QY	607	ttcaatccgggacttccatgcagacggccggccgtccctgcagatccctgattcttcgggtgcacgaac	666
Db	671	cttgcccgaggtctataccacgaagctcgcaagctgtaggctctctgctcccaaatgcagatc	730
QY	667	atcttcggagctgaccatccctctccatcacagggcccaaggttccatcgctcttaactc	726
Db	731	gtgtttgacctcccatctcgagacagccactccggccactctccctcaatctatacaagtctg	790
QY	727	gacatttgcataaaaaccccccacccaacaaatccgctctgactctactctgctctcaacgc	786
Db	791	gaggctcgcgcgtgaaagcgtgcctcccaagacaaggtctgcacacgggtgcctgcgcagctg	850
QY	787	ggctgctctccgcgtgcgtgtagaagagcttcaatgctcgcgtataatgcacacagaatctgcctc	846
Db	851	ggggtggtgcgtcgtggtgtagaagctgtgtaatgcacaagctgcagagacagctgcacatg	910
QY	847	cccatcccttaagagatgattgtgtgtgtgtgtgcacaacgctatctcccgagggtcgtgaag	906
Db	911	ccgatacccgaggagctgcgtctacgctcatatgcgtgcacagacagatctccatgtgcatgtgt	970
QY	907	atgcccaaaaaggtatcccaatgcagatctgtgtgggaaatcccaacgcgggtgtcccaacccg	966
Db	971	ctaaagacacagatttggtgtatgattgtcgtgtggaaataatccctgcagagggtcgtgtgccca	1030
QY	967	gtctgcgcgtgtgtctcaacaggtggaagagacatgatagcacagacctctccctagacatc	1026
Db	1031	gtggtccccaacaacacccagctgtttctcaagaactctgtgtgagagacctctcaacatcgctgtg	1090
QY	1027	gtgagctcagctacccaacccgtgcataatgggtgcggagacctgtgcacaagaagacggtcagac	1086
Db	1091	gtctggttttgcatatgcatcactcactcgtgggaagaaactctgcctgagggtgcacggctacccg	1150
QY	1087	gtggagtctgcgaacaggaagatgatacgtcctcgcgtcagcaactctgttgctctctcttc	1146
Db	1151	gtgcacagaacaacagagctgtgtgtcccttggtgcctcagtaactatctgcagagacatcttc	1210
QY	1147	aaaatctcatgtcatcttctgtgtgcgtcttcctgcacccctgcgcgtgtagatgtagctgtgaaag	1206
Db	1211	cagtgcctcccggtgaattgtgtctctcatgcacgcggagcctgtatccagagagaccccgaggcc	1270
QY	1207	aaatcccaagctggtgcacagccctgtgtgtgtctctgtgtgtatgcataccatgctgtgtccgt	1266
Db	1271	aaactgcaggtgtgtgtgtaggcacatctctctctcttcatccctcatccatctgttcaaaact	1330
QY	1267	gggtatctatctgtatctctctccctaaagtctgtgtctatggagccctatctgctgttcaatctc	1326
Db	1331	ggggaaactcttccatctgcactgtgcacaaagcgtgtctctgtgcacacatataatgtgtgaacctg	1390
QY	1327	aagaaatccctcccaagacacccaacccaccccttaactacgtgtggagagagagacagatgtgac	1386
Db	1391	aaggagcatgctgagtagcagctccagcgagacatgagctccctctgtgaaagccaactgcggcgatc	1450
QY	1387	tgctgcatctgtgtgtagttagctctctctctctctctctctctccacagctgcctbatgtgtgt	1446
Db	1451	ctgtcttatctgtgtgtgtaactcttcaacgcccacacatcttctgaacctgtgacctgtgacttgg	1510
QY	1447	gcaatgtggtgtgtgcctctctccctccgcgtgcgtgtgtgtgtgtcttcggaacacagatccccaatc	1506
Db	1511	gttgtgtgtgtgtgcatactctccctccgcgtgcgtgtgtgtgtgtgtcttcggaacacagatccccaatc	1570
QY	1507	tatgcacagtcaccaagctgcacatgcacatgcacattatgtatgcacaaagacatataatag	1566
Db	1571	ctgtctctggggcaggtgcgtgcacagacagatatttaagaagatgtgtgcagagtatactcagag	1630
QY	1567	gccacagatatccaggggtataaatcatcatcagctgctccctctctactattgtgcacac	1626
Db	1631	gccacagaaagctccgggtgtgtgaagctctccgctctctctgcgcacacggttatcttggcaat	1690

Oy	1627	tcagagatcttcag	1640
Db	1691	gcctgattctaacg	1704
RESULT 2			
ID	AAI62799		
	AAI62799 standard; cDNA: 807 BP.		
XX	AAI62799;		
XX			
DT	22-OCT-2001	(first entry)	
DE			
XX	Human CDNA SEQ ID NO 58.		
KW	Human; nontropic; neuroprotective; cytosolic; dermatological; virucide;		
KW	immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;		
KW	antipneumococcal; antischling; antlanemic; antarrhrilic; cancer;		
KW	antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;		
KW	antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;		
KW	antiparasitic; cardiant; immune disorder; cardiovascular disorder;		
KW	neurological disease; infection; nephrotropic; gene therapy; vaccine;		
ss.			
OS	Homo sapiens.		
XX			
PN	WO20015449-A1.		
XX			
PD	02-AUG-2001.		
XX			
PF	17-JAN-2001; 2001WO-USO1346.		
XX			
PR	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180628.		
PR	19-MAY-2000; 2000US-0205515.		
PR	07-JUL-2000; 2000US-0216880.		
PR	14-JUL-2000; 2000US-0218290.		
PR	14-AUG-2000; 2000US-0225447.		
PR	01-SEP-2000; 2000US-0229343.		
PR	06-SEP-2000; 2000US-0230437.		
PR	25-SEP-2000; 2000US-0234997.		
PR	29-SEP-2000; 2000US-0236367.		
PR	13-OCT-2000; 2000US-0239937.		
PR	08-NOV-2000; 2000US-0246476.		
PR	08-NOV-2000; 2000US-0246477.		
PR	08-NOV-2000; 2000US-0246526.		
PR	08-NOV-2000; 2000US-0246528.		
PR	17-NOV-2000; 2000US-0249210.		
PR	17-NOV-2000; 2000US-0249214.		
PR	17-NOV-2000; 2000US-0249265.		
PR	01-DEC-2000; 2000US-0250160.		
PR	01-DEC-2000; 2000US-0250391.		
PR	05-DEC-2000; 2000US-0251030.		
PR	05-DEC-2000; 2000US-0251988.		
PR	05-DEC-2000; 2000US-0256719.		
PR	06-DEC-2000; 2000US-0251479.		
PR	08-DEC-2000; 2000US-0251988.		
PR	08-DEC-2000; 2000US-0251990.		
PR	11-DEC-2000; 2000US-0254097.		
PA			
XX	(HUMA -) HUMAN GENOME SCI INC.		
PI	Rosen CA, Barash SC, Ruben SM;		
XX			
DR	WPI; 2001-476225/51.		
DR	P-P5DB; AAM42394.		
XX			
PT	Novel plasma membrane associated proteins useful for diagnosing,		
PT	treating, preventing and/or prognosing disorders related to the		
PT	proteins, including cancer, immune response and neuronal disorders		
-			

XX	Claim 1; SEQ ID NO 58; 532pp + Sequence Listing: English.
PX	
XX	The invention relates to novel genes (AA162752-AA162961) and proteins
CC	(AAM42347-AAM42415) useful for preventing, treating or ameliorating
CC	medical conditions e.g. by protein or gene therapy. The genes are
CC	isolated from a range of human tissues disclosed in the specification.
CC	The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC	in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC	and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC	marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC	(b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC	haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC	disease, multiple sclerosis, Rheumatoid arthritis and ulcerative
CC	colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC	(d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC	epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC	and parasitic infections.
CC	Note: The sequence data for this patent did not form part of the
CC	printed Specification, but was obtained in electronic format directly
CC	from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
CC	
XX	Sequence 807 BP; 181 A; 233 C; 194 G; 190 T; 9 other:
SQ	
Query Match	15.7%; Score 355.2; DB 22; Length 807;
Best Local Similarity	98.3%; Pred. No. 8.5e-79;
Matches 357; Conservative	1; Mismatches 5; Indels 0; Gaps 0;
OY	312 ctactctctcttccccctccctgaacctactcttcctgggagggtgtcaccaagatgt 371
Db	18 ctgcagagtctctctcccctcccttgaccactctctcmtggygggtgtacccaagatgt 77
OY	372 gccaggtaccttgcgtatcagaacctgtgtggttaacatctgtcgaactgtgcccc 431
Db	78 gccaggtaccttgcgtatcagaacctgtgtggttaacatctgtcgaactgtgcccc 137
OY	432 agagttcgaaattcccggtcttcaacaatgccaaccaatgtagagctatgtgacacagcagc 491
Db	138 agagttcgaaattcccggtcttcaacaatgccaaccaatgtagagctatgtgacacagcagc 197
OY	492 catgtagagcttgagaggtctgcacgttgtcaactacgcctagacctgtcctaacgcatatcca 551
Db	198 catgtagagcttgagaggtctgcacgttgtcaactacgcctagacctgtcctaacgcatatcca 257
OY	552 gatgggtccgggtccatcatgttggctttgtggccatctacacctccgaagtccat 611
Db	256 gatgggtccgggtccatcatgttggctttgtggccatctacacctccgaagtccat 317
OY	612 ccggggtcatgatcagcgccgcgacctgagatactctgatttcgggtctcaagtacatctt 671
Db	318 ccggggtcatgatcagcgccgcgacctgagatactctgatttcgggtctcaagtacatctt 377
OY	672 cgg 674
Db	378 cgg 380
RESULT 3	
AAC03684	
ID AAC03684 standard; CDNA; 443 BP.	
XX	AAC03684;
AC	
XX	06-OCT--2000 (first entry)
DT	
XX	
DE	Human secreted protein 5' EST, SEQ ID NO: 3682.
XX	
KW	Human; 5' EST; expressed sequence tag; secreted protein; CDNA isolation;
XX	gene therapy; chromosome mapping; ss.
OS	Homo sapiens.
XX	

PN EPI033401-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 21-FEB-2000; 2000EP-0200610.
 XX
 PR 26-FEB-1999; 990S-0122487.
 XX
 PA (GSEST) GENSEST.
 XX
 PI Dumas MLine Edwards J, Duclert A, Giordano J;
 XX
 DR WPI: 2000-500381/45.
 XX
 DR P-PSDB; AAG03678.
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 1; SEQ ID 3682; 71bp + CD-ROM; English.
 XX
 CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. An ORF has been identified within the
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
 CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.
 XX
 SQ Sequence 443 BP; 114 A; 123 C; 103 G; 96 T; 7 other;

Query Match 14.9%; Score 337.6; DB 21; Length 443;
 Best Local Similarity 98.9%; Pred. No. 1.6e-74;
 Matches 348; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

QY 1313 tgcgtcgaatctcaagaactccctcaagaactcaagccctactactcgttgagaga 1372
 DB 1 tgcgtcgaatctcaagaactccctcaagaactcaagccctactactcgttgagaga 60
 QY 1373 agagcaagctggaactgttcacatctgtgtagtgagcttcctcctcctcctccca-gc 1431
 DB 61 agagcaagctggaactgttcacatctgtgtagtgagcttcctcctcctcctcccaargc 120
 QY 1432 ctgcacctatgtgtgagcagtggtgtcgcctctccgcctcgtcgtgcttccagact 1491
 DB 121 ctgcacctatgtgtgagcagtggtgtcgcctctccgcctcgtcgtgcttccagact 180
 QY 1432 cagtttcaaatgactatgactgtgcccagatcagatgagacatgacattatgtgaatcc 1551
 DB 181 cagtttcaaatgactatgactgtgcccagatcagatgagacatgacattatgtgaatcc 240
 QY 1552 aagacctataataggccagagatatccagggtatataatatacagactgctccct 1611
 DB 241 aagacctataataggccagagatatccagggtatataatatacagactgctccct 300
 QY 1612 ctctacttgcgaactcagagatcttcaggcgaagaagtcacatgcgaagactg 1663
 DB 301 ctctacttgcgaactcagagatcttcaggcgaagaagtcacatgcgaagactg 352

RESULT 4
 AA071399
 ID AA071399 standard; CDNA: 2882 BP.
 XX
 AC AA071399;

XX
 DT 01-APR-1995 (first entry)
 XX
 DE Down-regulated in adenoma DNA gene.
 XX
 KW DNA; down-regulated in adenoma; tumor suppressor; colon adenoma;
 KW colon adenocarcinoma; diagnostic; therapeutic; ds.
 XX
 OS Homo sapiens.
 XX
 FT Key Location/Qualifiers
 FT CDS 185..2479
 FT /*tag= a
 XX
 PN W09420616-A.
 XX
 PD 15-SEP-1994.
 XX
 PF 04-MAR-1994; 94WO-US01860.
 XX
 PR 05-MAR-1993; 93US-0026045.
 XX
 PA (SCHW/) SCHWETNEST C W.
 PA (USSH) US DEPT HEALTH-& HUMAN SERVICES.
 XX
 PI Papas JS, Schweinfest CW;
 XX
 DR WPI: 1994-303023/37.
 DR P-PSDB; AAR60568.
 XX
 PT Isolated polypeptide that is down regulated in colon
 PT adenocarcinomas and adenomas - is used as an indicator of tissue
 PT abnormality.
 XX
 PS Claim 5; Page 21-25; 41pp; English.
 XX
 CC DNA probes may be synthesized according to the DNA sequence and
 CC used to screen colon tissue samples for DNA mRNA. Absence of DNA
 CC mRNA, and consequently DNA tumor suppressor protein, indicates
 CC tissue abnormality. The cDNA may have therapeutic activity.
 XX
 SQ Sequence 2882 BP; 839 A; 579 C; 595 G; 868 T; 1 other;

Query Match 14.4%; Score 326.2; DB 15; Length 2882;
 Best Local Similarity 51.7%; Pred. No. 2.4e-71;
 Matches 798; Conservative 0; Mismatches 733; Indels 12; Gaps 2;

QY 124 agatgttctcagccaagatacaagctgtgtgttgagctgtcctcgtctcctcgtg 183
 DB 314 agctgttcccccaaaaaggccaagataatgttcctctcttcttcccatagacatttg 373
 QY 184 ctccccaagtaagaatgaagatacatcatcctcgtcactgtcgtgtgactcagcg 243
 DB 374 ttgcagcagcaccggtcttaagaatgattgtgtcagtgatattgtttctgtatccagaca 433
 QY 244 ggaatccatccaggtccacaaagacatgcatctgtcctcgtgcgaacctctcgtcagc 303
 DB 434 gggattgtgcccgtactataaagatttagattgtcctcgtcgtgtcgaattcccccagc 493
 QY 304 aatgacctactcctcctcttccctcctcgtacactacttctccgtgggggtttcac 363
 DB 494 tatgggttgatgacatccttcttccagcgaataatcacttctctcgtcgaattccaga 553
 QY 364 cagatgtgtccaggtacacttgcgtttacagcatcctgtgtg-----taaca 412
 DB 554 caacatccgtgtgtccgttccgtatctcgtatgatagtgtgtgactgagattccagga 613
 QY 413 tctgtcgaagctggccccaagatcgaaatccaggtcttcaacaatgcc-accaatgag 471
 DB 614 gcagttccaagaagctcccgatcgcgaatgacactacttgggtgtgtcgaataacacgcg 673
 QY 472 agctatgtgacacagacagatgagagctgagaggtgtgacgtgtcagctagctagcc 531

Db	674	aataatcttctactactcgtgtagcggaggggtgaggggtggcgcgggccgcatcagtcaca	733
Oy	532	tgctcaacgcgcacatccatccagatvgtgtctggtcctcaatgcattgtgcttgtgcac	591
Db	734	gtgtcttcttgaaatcacaccagttgtgtcttggattcttcggatvgtatttgaatgata	793
Oy	592	tactctccagatcccttcacccgggctctcaatgaagcgccggcccttgagatccctgatt	651
Db	794	tactgtctcgtagctccctcacacagttgcttccactacgtcgtcgtctgttcaattttgttt	853
Oy	652	tcggtgtcacaagtacacatcttcgtgagctgcacatccctccctacacagggcccaagttccac	711
Db	854	tcaccaactcaaatcatcttttctcagtttgacgtcccgctcacacatgcattccagtttcaatt	913
Oy	712	gtcttaaccttcaatlgacatttggcaaaaacgtcccccacacacatgcctcgtctac	771
Db	914	ttcaaaagtaactactactctgtatcttccacaatagagaagaaactaatttgcagacctgtg	973
Oy	772	ttcgccttcacagcgggtgtgcctctccctgtgtgtgtgtaagagctcattcgtccatag	831
Db	974	acagctctgtatgtgccttcttgtgtatcatcattgttaagaataataatcagcgcttcaaa	1033
Oy	832	cacaaagattcgtctcccatccctacacagatvattgtgtgtgtgtgtgtaaacagctac	891
Db	1034	gacaaactctcagtcagcccatctccatctgaattcatattatccgttatcttcacaggtgta	1093
Oy	892	tcggggggtcgtgaagatggtcccaaaagatcacatcagatcgtgtggagaaatccacgc	951
Db	1094	tcctacggcgtgtgactttaaaaacaggttaaaagtggtcgtgtgtgtgtggagatgaatcct	1155
Oy	952	gggttccccaaccccggtgtcgctgtgtgtctcagltgaaagacatgataagcacagcc	1011
Db	1154	ggatttcagcccccattacacactgacgtgtgagacttccaaaacacgtaagagatg	1215
Oy	1012	ttcccccctagccatcggtgagctgcatacgtatcaactgtgcatgtggccggaccctgtgcac	1071
Db	1214	ttcgcatcgtcaatvgttgcattgttgcacgtgtgccttccatggttgcagcgctacatccctc	1273
Oy	1072	aagcacgcgtacgaacgttgcatactgaaacagaagaatgatcgtctccgtcgtccagcaatcc	1131
Db	1274	aaatacagattacacacttgaatgtgcacatcagagattaatagccttggactgttgataata	1333
Oy	1132	tttggcccttcttaaatatcaatgtcatcttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1191
Db	1334	gtctgtggaagatcacaagagattgtcgtggaagtacgtgcctctccagatccagatftcag	1393
Oy	1192	gatggagctgtaggaagaataatcccaagctgtgcagcgtctgtgtgtgtgtgtgtgtgtgtgt	1251
Db	1394	gagagcacagagagcaaaaacacacagatgtgcgtgggtctaatgtgtgtgtgtgtgtgtgtgt	1453
Oy	1252	acacatcgtgtcctgtggagatctatctgtatccctccctcaatgctgtgtgtgaagccctg	1311
Db	1454	gtcgtctcagcatltgtgattcttccctgtgcgctctcaaaaagcttccgttctctgtgcattta	1511
Oy	1312	atgcgtctcaatctccaaagaaactccctcaagaacactacacgaccctcactactgtgtggag	1371
Db	1514	gcatttggaaacttaaaagggaatgcgtgtagtgcattgtcgtgaatgaagcagatgtgtgcga	1573
Oy	1372	aagagcaagctgtgacttctgcactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1431
Db	1574	aagagcaaatatgtgtttaaatttgatcatgaacttcaacttcaaccatttccctgtgcga	1633
Oy	1432	ctgcacctatgtgtgtgcagctgt	1491
Db	1634	ctcgtgtttagtcctgtggaagcttagtgtgtgcatcttcaactgttaaccatctgtgtgtcaggac	1693
Oy	1492	cagttctgaaatgtgtctatgcacgtgcccaagctatgtgacactgacattatgttgatccc	1551
Db	1694	caatttccaaaatgtcagcagcgcgtgtgtcaaatatgtggaagaacacacatctataagaataa	1753
Oy	1552	aagacctataataggtcccgagatcatccaggggattaaatcatcatcglatctgtccct	1611

Dd	1754	aagattatgatgatactggccagaaggagcgaaatttttcgaagtgcacatcctt	1813		
Oy	1612	ctctacttgccaactcagatcatcttcaggcaaaaagtcacg	1654		
Dd	1814	atctacttcgaacaatgttctttagcggaacttacg	1856		
<hr/>					
RESULT 5					
	AAx78074				
XX	AAx78074 standard; cDNA; 2792 BP.				
AC	AAx78074;				
XX					
Df	19-AUG-1999 (first entry)				
DE	Rat DTDST cDNA.				
XX					
KW	DTDST; human, expression vector; sulphate transporter; screening;				
RN	bone disease; cartilage disease; treatment; cell sulphate ion intake;				
KM	drug preparation; rat; ds.				
XX					
OS	Rattus sp.				
XX					
FH	Key	Location/Qualifiers			
CDS		368..2587			
FT	/tag= a				
FT	/product= "DTDST"				
XX					
PM	Jp1146790-A.				
PD					
XX	02-JUN-1999.				
XX					
PF	18-NOV-1997; 97JP-0335157.				
XX					
PR	18-NOV-1997; 97JP-0335157.				
XX					
PA	(SUMU) SUMITOMO SEIYAKU KK.				
DR	WPI; 1999-378999/32.				
XX					
PT	P-PSDB; AAY08929.				
XX					
XX	Sulfate transporter gene expression vector				
PS					
XX	Example 1; Page 16-19; 22pp; Japanese.				
CC	This invention describes the construction of a novel vector for sulfate				
CC	transporter expression containing a DNA sequence encoding a mammalian				
CC	sulfate transporter (expression product of the DTDST gene) and				
CC	containing no DNA sequence of 5' translation region of the mammalian				
CC	sulfate transporter gene. The invention also describes; (A) a method for				
CC	screening a human bone/cartilage disease treating agent including the				
CC	steps: (1) transforming an animal cell with the above vector; (2)				
CC	culturing the animal cell in the presence of a sample and (3) detecting				
CC	the increase in the sulfate ion intake to the cell; (B) a drug				
CC	preparation for the treatment of human bone/cartilage disease containing				
CC	the above vector as the active component. The sulfate transporter				
CC	gene-containing vector is high in expression efficiency. This sequence				
CC	encodes the rat DTDST protein used in the method of the invention.				
XX					
SO	Sequence 2792 BP; 694 A; 651 C; 661 G; 786 T; 0 other;				
<hr/>					
Query Match 10.7%; Score 242.2; DB 20; Length 2792;					
Best Local Similarity 49.3%; Pred. No. 2e-50; Indels 38; Gaps 4;					
Matches 775; Conservative 0; Mismatches 758;					
Oy	106	aaactcgcgaagctcagatgtctcctcagccaagatcaaacgctgtgttgagctg	165		
Dd	584	aagctcagaagaagtgccagtgtaatgcacaaccaaatcagaatatagatttgattt	643		
Oy	166	ctgctgtgtctctcagctcccacaagtaacaagattaagactacatctcgtacctg	225		
Dd	644	tctctgttcttaagtgctcccaaatatgatactcgaagaanaaacatttagtgacctg	703		

QY 226 ctcggtggaactcagcggggatccatccaggtccccaaggcattgtctcgcg 285
 Db 704 atgtctcgccgtatgttggttatattgttggcgcagtcatttctactcccggtg 763
 QY 286 gccaaactctcgcagtcataatggcctctactcctcttccctcctgactcttc 345
 Db 764 gctgcgaagaaactctactatgtctgtcacatcatcttltgcagcatattacttc 823
 QY 346 ttccctgggggtgttcacagaaatggtccaggtactttgcgttataccagctcggtg 405
 Db 824 cgtgttgcatactccgcacacatctctgtgggcaattttggaaactgtgccttaagat 883
 QY 406 ggt-----aacatctgtcagctggcccgagagtcgaaatcca-----gg 448
 Db 884 ggtgaggtatgttgaaccgagaaactaataagacctgcccgaacttgatactaatatct 943
 QY 449 tcttcaacaatgcaccaatagaactatgt-----ggacaagc 488
 Db 944 tccatagcaatgtttccaatgtatgtgtgtagtaaacatatacatataagacgactctgt 1003
 QY 489 agccatggaggtcagagagctgcagctgtcagctagctagcctgcctcaaccatct 548
 Db 1004 gacaaaagctgtatgcataatgaattggcagc-actgtgacatccaatgagctggagttta 1062
 QY 549 ccagatgggtctgggcttcaatgcagtttgcgttcttgcacatctacctccgagctct 608
 Db 1063 tcagtgacacatgggctcttctcaagtggccttgcctgtcgtctactcctcagatgcctt 1122
 QY 609 catccgggcttcaatgacggccggcctgcagatcccgatcttggtcgaatgacat 668
 Db 1123 gctgagcgggtgtgtcatctgtgtcctccaacatccatccatccagctcagagtaact 1182
 QY 669 ctccgagctgacacatccctcctcacaagccaggtccatcgtcttactcctcaatga 728
 Db 1183 cctggggcgtgagcctccctcccgagcaatgtgtgagctcagttcattactacatcgatcca 1242
 QY 729 catctgcaaaaactccccaacacaacatcgctcgtcactccttgcctcctcagcgg 788
 Db 1243 catcttcagaataatcatcaatagaaccaacatctgtgacccatccagcgttltgtct 1302
 QY 789 tgccttcctcgtgtgtgtgaagagatcaatgtcgtctcaatgacaaagatctgctcc 848
 Db 1303 cctgtgtccttgtgcacaacaaagaactcaagaatactccaagttccacccggcaco 1362
 QY 849 catccctacagagatgtgtgtgtgtgtgtgcacaagatctccggggcctgtaagat 908
 Db 1363 aattccaactgagctattgtcgtgtgtgtgcagcacatgtgctctcattttggaaact 1422
 QY 909 gcccaaaaagtacacatgcagatcgtgtggagaatccaaacggggtccccaacccggt 968
 Db 1423 caatggaattgaattcagatltgcgggcaaatcccaacgggttltatgcaccoca 1482
 QY 969 gtcggcctgtgtctcagctgtaagagacatgaagcaagccttccctcagcactgt 1028
 Db 1483 agcgccagagctggagcctcaatctcctaagtgtgtgtgtgtgtgtgtgtgtgtgtgt 1542
 QY 1029 gaggctacgtacaaactgtgtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1088
 Db 1543 tggtttgtctacactgtatacttctcgtatgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1602
 QY 1089 ggaattcgaaccagagatgactgcctcgcgtgcagcaacttcttgcctctctttaa 1148
 Db 1603 gaaagcaatcagaatgtatgtccattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1662
 QY 1149 aattcatgtcattgt 1208
 Db 1663 ctgcataactatgt 1722
 QY 1209 atcccaagtgccagcgt 1268
 Db 1723 gacccaactgtcgcgtatgtgaatccctcgttcttctgtgtgtgtgtgtgtgtgtgtgt 1782

QY 1269 gactatctgtatccctcctcctaagtcgtgtgtaggagccctgactgcgtgtcaatccaa 1328
 Db 1783 tccctattctattccctcccaaaaatgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1842
 QY 1329 gaactccctcaagcaatccacgacccctactaccctgtgtggaagaagaagctgtgactg 1388
 Db 1843 ggggtgactcttgaatttagaagacctgcgaagatgtgtgagctcagaagaatgagacac 1902
 QY 1389 tgcagctggtagtgaatctccctcctcctcctcctcctcctcctcctcctcctcctcctc 1448
 Db 1903 agttatctgt 1962
 QY 1449 agtgggtgtgcctctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1508
 Db 1963 cgttgggttcttcttcaatgttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2022
 QY 1509 tgcactggccaggtatgtgacacttacttattgtgaaatcccaagactataatgggc 1568
 Db 2023 actgctgtgttggaaagaatctgaatctcgatccatctccactcaagaagacct 2082
 QY 1569 ccaggaatccagggattaaatcaatcaagctgtccctcctcctcctcctcctcctcctc 1628
 Db 2083 tgggaatagctcagagcatcaaggtttccgcttcatagccctcctcctcctcctcctc 2142
 QY 1629 agagatcttca 1639
 Db 2143 agaatgtctta 2153
 RESULT 6
 AA11766
 ID AA11766 standard; DNA; 1465 BP.
 XX
 AC AA11766;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Probe #1699 for gene expression analysis in human cervical cell sample.
 XX
 KW Probe: human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157278-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00670.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 XX
 PR 26-MAY-2000; 2000US-0207456.
 XX
 PR 30-JUN-2000; 2000US-0608408.
 XX
 PR 03-AUG-2000; 2000US-0632366.
 XX
 PR 21-SEP-2000; 2000US-0234687.
 XX
 PR 27-SEP-2000; 2000US-0236359.
 XX
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488901/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human cervical epithelial cells -
 XX
 PS Claim 25; SEQ ID NO 1699; 487bp; English.
 XX
 CC The present invention relates to human single exon nucleic acid probes
 CC (SENP). The present sequence is one such probe. The SENPs are derived
 CC from human HeLa cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a

CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging
 CC of diseases of the cervix, notably cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 1465 BP; 389 A; 319 C; 284 G; 473 T; 0 other;

Query Match 8.7%; Score 196.6; DB 22; Length 1465;
 Best Local Similarity 48.1%; Pred. No. 3.5e-39;
 Matches 556; Conservative 0; Mismatches 599; Indels 0; Gaps 0;

QY 546 catccagatggtcgtggtctcatcaggttggcttggscatccatccctccagtc 605
 DB 111 cttccagtgagtaggtggtctcttcaagtggttggcttggscatccatccagtc 170
 QY 606 ctcaatccggggtctcaatgagcgccggtccagatcctgattcgtgtcaagta 665
 DB 171 ctgtcgtgagtgattgtcactgtgctccttcaactattctacatctcagccaagta 230
 QY 666 catctcgagatgacatccctccctctacacagggccaggggtccatccttcat 725
 DB 231 tctcttgggtccacactctccctcggaactaagtggtgggtccatccatccatccgtgat 290
 QY 726 tgacatttgcaaaaaccccccacacacacacacacacacacacacacacacacac 785
 DB 291 acatgtctcagaacacatccatcagaacacacacacacacacacacacacacacacac 350
 QY 786 cgtgtccttctgtgtgtgtgaaagagctcaatgtcgtacatgacacaaatcgtt 845
 DB 351 ccttgggtcttcttgcacacaaagaactcaatgacacacacacacacacacacacac 410
 QY 846 ccccatccctcagaagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 905
 DB 411 accgatctccatcgaactgt 470
 QY 906 gatgcccacaaagatcacaatcagatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 965
 DB 471 actaatgaaatataatctctagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 530
 QY 966 ggtgtcgt 1025
 DB 531 caaagatcagaatgaaacctaattcctagtggtgtgtgtgtgtgtgtgtgtgtgtgt 590
 QY 1026 cgtgagcagatcacaacgt 1085
 DB 591 catgt 650
 QY 1086 cgtgagatcagaacagagatgagatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1145
 DB 651 agtcaagaacacacaggaatgt 710
 QY 1146 taaatcattcattgt 1205
 DB 711 ccactgttctactactgt 770
 QY 1206 aaatccaggt 1265
 DB 771 ccatctacagcttctgt 830
 QY 1266 gggagatcactgtatctctccctcaagtctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1325
 DB 831 agctccttctgtctatctccctcaaaaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 890
 QY 1326 caagaactcctcagaac 1385
 DB 891 acggggagccctctgtaaatatagagatctcccaaatgtgtgtgtgtgtgtgtgtgtgt 950
 QY 1386 ctgt 1445
 DB 951 taagttatctgt 1010

QY 1446 ggcagtggt 1505
 DB 1011 actgt 1070
 QY 1506 ctatcagctgt 1565
 DB 1071 ttcactgt 1130
 QY 1566 ggcacagatattccagggatattcaatcaccagctacgtctccctctactattgtgcca 1625
 DB 1131 ccttaagactaagcagacatcaaatgttccgttggtagcctctctactataataa 1190
 QY 1626 ctcaagatcttcaggcaaaagtcacgcacagactgtctcctcgcagagatgacga 1685
 DB 1191 caaagaatgtcttaattcgtgttatacaacaactgtcaacccaatctataaaggt 1250
 QY 1686 ggaacttgagaatgc 1700
 DB 1251 ggccttgagaagaagc 1265

RESULT 7

AA133076
 ID AA133076 standard; DNA; 1465 BP.

XX AA133076;

DE 17-OCT-2001 (first entry)

DE Probe #1762 used to measure gene expression in human placenta sample;

KW Probe; microarray; human; placenta; antenatal diagnosis;

KW genetic disorder; ss.

OS Homo sapiens.

PN WO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632365.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488897/53.

PT Human genome-derived single exon nucleic acid probes useful for

PS analyzing gene expression in human placenta -

XX Claim 25; SEQ ID No 1762; 654bp; English.

XX The present invention relates to single exon nucleic acid probes (SENP).

CC The present sequence is one such probe. The probes are useful for

CC producing a microarray for predicting, measuring and displaying gene

CC expression in samples derived from human placenta. The probes are useful

CC for antenatal diagnosis of human genetic disorders.

XX Sequence 1465 BP; 389 A; 319 C; 284 G; 473 T; 0 other;

Query Match 8.7%; Score 196.6; DB 22; Length 1465;
 Best Local Similarity 48.1%; Pred. No. 3.5e-39;

Db 111 ctccaggtgagatggtctcttcaagtggttcttctcttcaactctcaatgagtc 170
 QY 606 ctcaatccggggtctcaatgagcgccggtctcagatccatcttctggtcgaagta 665
 Db 171 ctgagtgagtggttctgagctgtgctctctctcaatcttcaatctcaagta 230
 QY 666 catcttcgagtgacatccctccctcaacagccaggttcacatgcttaccatc 725
 Db 231 tctcttggtctcaactctcctcgactaagtggttggtcctcaatcactatgag 290
 QY 726 tgacattgcaaaacctcccccaacaatctcctcgatctcttctgctcactag 785
 Db 291 acatgtctcaagaacatccataagacaatctctgtagctatccaccagcttg 350
 QY 786 cgtgctctctctggttggtggaagagctcaatgctcgtctacatgccaagatc 845
 Db 351 ccttctggtctcttgcgaacccaagactcaatgaaacactcaatccaagctaa 410
 QY 846 ccccatccctcaagagatggttggttggttggtggaacatctcccggtgttaa 905
 Db 411 accgatccctatgaactgtgtgtgtgtagcagccacatlagcctctcatctg 470
 QY 906 gatgccccaaaagatatacatatgcagatcgtggtggaatccaaagcggtcc 965
 Db 471 actacatgaanaatataatctagatgtgtgtagacataatccacgtgtttag 530
 QY 966 ggtgctgctgtgtgtctcacagtgaagagatagagacagacgtctccctag 1025
 Db 531 caaagtctccggaatggaacctaattcctagtggtggtgtagatgcaatagct 590
 QY 1026 cgtgagctacgtcaacacgtggtatggtggtggtggtggtggtggtggtg 1085
 Db 591 catgtgttctgctacatcgtatcactcttcgagatgttgcgaagaacatgt 650
 QY 1086 cgtggtatggaacaggaagatgagtcgtcgtcgtcgtcgtcgtcgtcgtc 1145
 Db 651 agtcaaaagcaaacgaagaatgtagtgcactgtgtgtgtgtgtgtgtgtgtgt 710
 QY 1146 taaatcatgtcatcttggtggtggtggtggtggtggtggtggtggtggtg 1205
 Db 711 ccaactgtttactactagtgagctcttgcaagaacatgtgttaagaatcaac 770
 QY 1206 aaaaatccaggtggtcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1265
 Db 771 ccatctcagcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 830
 QY 1266 ggggatcatctgtatctctcccttaagtcgtgtgtagagccctgagctgt 1325
 Db 831 agctccttctgtctatctccctcaaaaagtgtccctgtgtgtgtgtgtgtgt 890
 QY 1326 caaagaatccctcaagaacactcacccactactactctggtggtggtggtg 1385
 Db 891 accgggggacctctgtaatttaggagatcttcccaaatgtgtgtgtgtgtgt 950
 QY 1386 ctgtgcatctggttagtgagcttctcctcctcctcctcctcctcctcctc 1445
 Db 951 tacagttatctggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1010
 QY 1446 ggcaggtggt 1505
 Db 1011 actgt 1070
 QY 1506 ctatgactggtccaggtcagacatgacatgacatgacatgacatgacatgac 1565
 Db 1071 ttaactgt 1130
 QY 1566 ggcagagatatacaggtggtggtggtggtggtggtggtggtggtggtg 1625
 Db 1131 ccttcagagtaagcaggtcagacatggtgtgtgtgtgtgtgtgtgtgtgtgt 1190
 QY 1626 ctgagagatcttgagcaaaagtcatcgtccaagactgtccctcgtcaggg 1685
 Db 1191 caaagaatgcttataatctgtcttatacaaaaacgttcaacccaatctta 1250

QY 1686 ggaacttgagaatgac 1700
 Db 1251 gacttgagaagaagc 1265

 RESULT 9
 AAI20978
 ID AAI20978 standard; DNA; 1521 BP.
 AC AAI20978;
 DT 12-OCT-2001 (first entry)
 DE Probe #10911 for gene expression analysis in human cervical cell sample.
 DE XX
 KW Probe: human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer; ss.
 OS Homo sapiens.
 PN M0200157278-A2.
 PD 09-AUG-2001.
 XX 30-JAN-2001; 2001WO-US00670.
 PF 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI; 2001-488901/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human cervical epithelial cells -
 XX
 PS Claim 25; SEQ ID No 10911; 487bp; English.
 XX
 CC The present invention relates to human single exon nucleic acid probes
 CC (SNP). The present sequence is one such probe. The SNPs are derived
 CC from human HeLa cells. The SNPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging
 CC of diseases of the cervix, notably cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WPI
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 1521 BP; 412 A; 325 C; 309 G; 475 T; 0 other;

 Query Match 8.5%; Score 192.2; DB 22; Length 1521;
 Best Local Similarity 48.0%; Pred. No. 4.4e-38;
 Matches 548; Conservative 0; Mismatches 593; Indels 0; Gaps 0;

 QY 560 tgggttaagcagttggt 619
 Db 8 tgggtcttcttcaagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 67
 QY 620 tcatgagcgccggtcgtcagatctggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 679
 Db 68 tgtgacgtggtggtccctcctcaactatcttaactcctcaagcagatcttctgtg 127
 QY 680 ccatccctcctcaacagggccaggttccatcgtcttactcttcatgtacattgtg 739

[illegible]

AAI46225
ID AAI46225 standard; DNA; 1521 BP.
XX AAI46225; `
AC AAI46225;
XX AAI46225;
DT 17-OCT-2001 (first entry)
DE Probe #14911 used to measure gene expression in human placenta sample.
XX Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX Homo sapiens.
OS Homo sapiens.
PN WO200157272-A2.
PN WO200157272-A2.
PD 09-AUG-2001.
XX 09-AUG-2001.
PF 30-JAN-2001; 2001WO-US000663.
XX 30-JAN-2001; 2001WO-US000663.
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX MOLE-
PA MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI: 2001-488897/53.
DR WPI: 2001-488897/53.
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
PS Claim 25; SEQ ID NO 14911; 654bp; English.
XX The present invention relates to single exon nucleic acid probes (SENP).
CC CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
SQ Sequence 1521 BP; 412 A; 325 C; 309 G; 475 T; 0 other:
Query Match 8.5%; Score 192.2; DB 22; Length 1521; Best Local Similarity 48.0%; Pred. No. 4.4e-38; Matches 548; Conservative 0; Mismatches 593; Indels 0; Gaps 0;
Y 560 tgggcttcacgtgaagtttgcttgcgtcgtccatcacctcccgagcttcacgcgggct 619 D 8 tgggcttccttaaaatgggttgttcttcgtcatacctccagatgacttgtagtgat 67 Y 620 tcattgacgccgcgcgcgcgcgcgatctctggtggtcacaagtaccatcttcggactga 679 D 68 ttgtcaatcgtgtccttccttacatactctaacatcccagaagaagatatcttcgtgcca 127 Y 680 ccatccccctctacacagaggccagggtccatcgctcttaaccttcattgaacattgcaaaa 739 D 128 accttccttcgtgactaatgtgtggtgcctaactacatcaactcgtgataatgcttcagaa 187 Y 740 acctcccccaaccaatcgctcgtcatacttccttcgtccatcaagcgggtgcttcctg 799 D 188 acatcatagaagcaaatcctctgtgacttatcatcacagcctlttgccclttggtctc 247 Y 800 tgcctgtgaagaagtcaatgtctgtcactacagcaaatgcttcctccatccctacag 859 D 248 tggcaaccaaagaacatgaacacttcaaaatccaagccttaaggcacccagatctccatly 307 Y 860 agatgaattgtgtgtgtgtgcaaaagcatalctccgggggctgtlaagatgcccanaagt 919

! |||
Db 482 actatcg 489

Search completed: April 25, 2002, 14:41:58
Job time: 6000 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2002, 11:32:31 ; Search time 3196.3 Seconds

(without alignments)
11674.943 Million cell updates/sec

Title: us-09-749-589-1

Perfect score: 2262

Sequence: 1 atgagccagccagccagccgcg.....agaccctgacgcctgtga 2262

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rnd:*
33: em_htgo_hum:*
34: em_htgo_inv:*
35: em_htg_rnd:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	392.4	17.3	2855	10	AF337809
2	385.6	17.0	2589	9	AF288410
3	385.6	17.0	2748	9	AF279265
4	364.4	16.1	2629	10	AF136751
5	358	15.8	4113	10	AF230376
6	355.8	15.7	2612	10	AY032863
7	343.4	15.2	2853	10	AF167412
8	339.4	15.0	3086	10	AF167411
9	337.2	14.9	2235	10	AY024359
10	334.6	14.8	1389	4	AF314619
11	332.8	14.7	2235	10	RMO30372
12	332.4	14.7	1634	10	AF314820
13	328.2	14.5	4930	9	AF030880
14	326.2	14.4	2881	9	HUMDRA
15	326.2	14.4	2881	11	G18336
16	326.2	14.4	2882	6	AR052312
17	326.2	14.4	2882	6	AR144265
18	326.2	14.4	2882	6	I28252
19	289	12.8	62357	2	AL451008
20	289	12.8	62357	2	AL360009
21	259.6	11.5	2205	4	OAR18558
22	242.2	10.7	2792	6	E26555
23	242.2	10.7	2877	10	D82883
24	230.2	10.2	3076	9	AF297659
25	225.8	10.0	4741	4	B07223615
26	223	9.9	3711	10	RATSAT1A
27	222.8	9.8	3732	10	MUSROB
28	213	9.4	2832	9	HSU14528
29	211.2	9.3	1259	10	AF315652
30	204.4	9.0	178189	2	AC019103
31	196.6	8.7	175550	9	AC008427
32	196.6	8.7	235808	2	AC011406
33	190.2	8.4	165443	2	AC079939
34	182.8	8.1	185081	2	AC079937
35	143.2	6.3	185081	2	AC079937
36	119.4	5.3	32037	2	AC018283
37	109.8	4.9	40197	3	CBRG33P21
38	98.4	4.4	10208	1	AE004592
39	85.6	3.8	3008	9	HSMB00532
40	76.6	3.4	39874	3	CEK287
41	73.8	3.3	2403	9	AK023366
42	71.6	3.2	2515	8	SSSULPTRP
43	71.4	3.2	88326	9	AC005923
44	71.4	3.2	204513	2	AC024179
45	71	3.1	89203	2	AC010691

ALIGNMENTS

RESULT 1
LOCUS AF337809 2855 bp mRNA
DEFINITION Rattus norvegicus down-regulated in adenoma protein mRNA, complete cds.
ACCESSION AF337809
VERSION AF337809.1 GI:15077506
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 2855)
Ye,H.J., Binder,H.J. and Rajendran,V.M.
TITLE Molecular cloning and characterization of down-regulated in adenoma (DRA) mRNA from rat colon
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2855)
Ye,H.J., Binder,H.J. and Rajendran,V.M.

TITLE Direct Submission
JOURNAL Submitted (17-JAN-2001) Section of Digestive Diseases, Yale
University School of Medicine, 333 Cedar Street-1080 LMP, New
Haven, CT 06520-8019, USAFEATURES
source location/Qualifiers
1..2855CDS
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
191..2464
/note="DRA: membranous transporter"
/product="down-regulated in adenoma protein"
/protein_id="AAK8321.1"
/db_xref="GI:15077507"
/translation="MIEAIQNVVAVRPVSTKAGEEKKTYGHHKFLDLKCCS
CSQAKKIALSLPEIASMLPAKIKEMILSDIVSGIVGVALVLOGIALFALLVMP
AVGLAFEPVTFPELGSRIHSYSPVSMGVVTRVASSSDPALSSSAE
NDSMIEEKVMAASTVLSGIQLLGLVLOISFVITYSESLISGFTAAHIVLSQ
LKEMLQTVPAHSDPFSIFKLVESVSOIKNTADLVTSVILVAVPEKINRYS
SKLPVPIELIMTVIANGISYCNFEDRFVAVVGNMSIFOPITPSEVEFODTIG
DCFGIAGFAVAFSVASYSLKDYPIDGNOELIALGSNIFGAFKGFASLTISR
SGVSTGKTOVAGLSAVILVIVAGFLOPLOSVALAAGNKGMLQFAE
IGRLMKKXKDLIMTIFPIAIVIGLIGLAASAROLLTVRTPORPKSTLANC
RSNITKNNKNADYVEPBGVKTFRCPSPITYRNIFPKOKLIDAYGFNPLIKRKM
ALKTKRLQKGLDVTPKFTCTSDGFKDSDELDNNOIELDPIINTDLPFEDIDW
NADPLNITIPKISLHSLIDFSAVSLDISMRRLTILQEFIRIKVDVYIVGTDDW
FIDKLARCEFPDEVDYDSIFFLIHDALHIMWKDYSTSKFNSQERKEDFTINT
NGLRNRCQVPEVETK"BASE COUNT 790 a 640 c 620 g 805 t
ORIGINQuery Match 17.3%; Score 392.4; DB 10; Length 2855;
Best Local Similarity 53.7%; Pred. No. 7.2e-73;
Matches 885; Conservative 0; Mismatches 751; Indels 12; Gaps 3;

QY 22 taagtggtgacagagccgcatatccctaccctcttgcagatgaatttgagaagaag 81
DB 215 TATGTGTGGGACAGCCGGTGATTTCCACAAAGCTTTCGGAGAGCATTTAAGACAA 274
QY 82 gaccgacacatacccaatgagagaacttcgaatg----ccttcagatcttcacagcc 138
DB 275 TACGGACATCACAGACATTCCTCGATCATCTCAAGGGGTGTTGACGTCTCTACAA 334
QY 139 aagatcaaaagctgtggttgctgagcctgctgctccttcctgctccccaagaaga 198
DB 335 AAGGCCAAGAAATTCCTGCTTGTCCCATAGCATCTTGTCGACAGATACAG 394
QY 199 attaaagactacatcttccctgacccctgcgttgagactaagcggggagatccatccagtc 258
DB 395 ATAAAGGAATGGCTTCTCAGTACATCGTCTCGGCATGACACTGGGCTGTGGCTGTG 454
QY 259 ccacaagagatgacatttgcctgctgagcccaacttcgcgaagtaagtgcctctactc 318
DB 455 CTGCAAGATTAGCATTTCTCTGCTGCTGCAACATCCCTCTGCTATGAGGTTGTAACGA 514
QY 319 tcccttcctccctccctgacactctctcctgaggggagtgctcaaccagaatggctcagg 378
DB 515 GCCTTTTCCAGTATACACTACTCTCTTCTGGGACATCATGACACATATCTGGGCT 574
QY 379 acccttgcgtatcagatctcctggtggttaacatctgctcgcagctggccccaagagtc 438
DB 575 CCGTTTTCAGTTCTGAGCATGTATGGGGAGTTGATTAAGAAGTGGCTCGGGCTCC 634
QY 439 aaattccaggtcttcaacatgccaaccaatgagagctatgttgacaacagcagccatggag 498
DB 635 GAT-----ACTTCTCAAGCAATTGTCTCA-----AGCTCGCGAAGAAATGATTCAATGATA 685
QY 499 gctgagaagctgcacgtgtcagctagcagctgcctcaccgcacatcaccagatggc 558
DB 686 GAGGAGAAGGTAATGATGGCCGACATCGGTGATCTTCTTGGAATCATTTCAAGTTGCCTC 745
QY 559 ctggagctcaatgacgttgcttgcttgagccatcactcctcagatccttccatccggggc 618

DB 746 CTGGGGGCTTCGAGATAGGCTTTGTGTGATATACCTATCCGAGTCCCTAATACAGCGGC 805
QY 619 ttaatgagagccggcggcgtcgaagatccgatttggtgctgaagatcttcggagc 678
DB 806 TTTCACACGCGCTGCTCCATTCAGTGTGGTTTCCATCGAATTCATGCTACAGCTG 865
QY 679 accatccctctcacagaagcccaagggtccatcgtcttacttcaacttgaatltgacaat 738
DB 866 ACGGTCCAGACACAGCAGACCCATTCCTATATTAAGTCCGTGAGTCAGCTTCTCA 925
QY 739 aacctccccaacccaacatcgctcgtcactccttcgtctcaatcgaagggtccttcgt 798
DB 926 CAATTCACAGACGCAACATCCGACATCGAGCTGTCATCTGTGATATTCCTCGTGGCTCG 985
QY 799 ggtcgttgaaagagatgaatgctcgtctcaatgacaaagaatcgcttccccaatccata 858
DB 986 TTTGTCGTCAAGGAATTCACACAGCGCTACAGAGACGCTTCGGTGGCCATCCCATTC 1045
QY 859 gagatgattgtgtgtgtgtgcaaacagctatcccgagggtcgtlaagatgcccacaaag 918
DB 1046 GAACATCATGATGACTGTATTTGCAACAGGCATATCTATGGCTGTAACCTTGAACAGAG 1105
QY 919 tatcacatgcagatcgttggagaaatccaaacggaggtccccaacccgggtgcgctgtg 978
DB 1106 TTTGGGCGTGGCGGTGGGAAACATGAGCTTGGGTTCAACCCCTATTACCCCACT 1165
QY 979 gttccacgttggaagagacatgataagacacagctcttccctcaatcgttgagcagtc 1038
DB 1166 GTGGAAGTTTCCAGACACCATAGAGACTGCTTTGGCATTCGATTTGCGCTTTGGC 1225
QY 1039 atcaacctgtgctatggccggagccctcgtccacaacagcagctgacgagtgatcgaac 1098
DB 1226 GTGGCTTTTACGATTGCTAGCTGATCTGATTCCTCAATGAGATTATTCATGAGCAAC 1285
QY 1099 caggagatgattgctctcgtgcgcgaacacttcttggtcctcttcaaatcaatgctc 1158
DB 1286 CAGGAGTTAATGCTCTGGGGGTGAGCAACATATTCATCTGAGCTTCAAGGATTTGCA 1345
QY 1159 atttgctgctgctctctcactcctgctgctgctgagctgagagaaatccccaagtg 1218
DB 1346 GGGAGCACAGCCCTTCCAGATACAGGGCTTCAGAGAGACAGCAGGAGCAAAACACAGTTC 1405
QY 1219 gccagcctgtgtgtctctggtggtgataccaatcagctgctgctcgtggagatactctg 1278
DB 1406 GCGGGCTTCTCTCGGCTGTATTCGATGAGTATGATGATGATGATGATGATGATGATG 1465
QY 1279 tatcctctccctcaagctctgtctgtagagccctgcatcgtcgtcaatctcaagaatccctc 1338
DB 1466 CAGCTCTACAAAGCTCTGCTGCGAGCTCTGGCGCTTGAAACTTGAAAGGAATGCTG 1525
QY 1339 aagcaatcacagacccccaactaactcctgtgaggaagaagaagctgagactgtgcatctg 1398
DB 1526 ATGCAAGTTTCCGAGATAGGACAGGCTGTGGAAGAAAGATTAATAGCATTTGTTAATCTGG 1585
QY 1399 gtagtgaagcttccctcctcctcctcctcctcagcctgcccatagtgtgagcagtggtgc 1458
DB 1586 ATCATGACCTTCATCTTCCATCGTCTGCGGCTGAGATGAGGCTGCGACAGCAAGTGTG 1645
QY 1459 gctctcctcgtctcgt 1518
DB 1646 GCATTTTCAGCTCTTAATGTTGTGTGTCAGGACCCCAATTCACCAATTCAGCAGCACTGGCT 1705
QY 1519 caggtcatgacacagacattatgtgaatcccaagacctaataatgaagcccaagatctc 1578
DB 1706 AATGTTTGAAGAGACAACTTCAAGAAATTAATAAATAATTAATGCTGACGCTTTATAGCA 1765
QY 1579 caggagatataaatcaatcaatcagctacgtctccctcctcacttctgccaactcagaagatctc 1638
DB 1766 GAAGAGGTAAATTTTTCAGATGTCGCTCTCAATCTACTTCTGCAACAATTCGTTCTTTT 1825
QY 1639 aggcacaaggtcatcgccaagactgtct 1666

	BASE COUNT	552 a	801 c	785 g	610 t
ORIGIN					

Query Match	Similarity	17.0%	Score 385.6	DB 9	Length 2748
Best Local	Similarity	54.2%	Prod No. 2e-71		
Matches 821	Conservative	0	Mismatches 669	Indels 24	Gaps 1
QY	127	tgctccagcccaagaatcaaaagctgtygtgtctgagctgctgctgctcctgctc	186		
DB	372	tgctccagccgctgcggggccctatgctctgctgctcgaacacctccgggttttgctggtta	431		
QY	187	cccaagttacaagattaaagaatcatatcctgacccctgctgctgtygtactcaggggga	246		
DB	432	ccccggtatccctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	491		
QY	247	tcacatcaggtcccaagaagcaatgcatctgtctgctgctgctgctgctgctgctgctg	306		
DB	492	atcatgacacttccgagggcttggctgacgcttccctgctgctgctgctgctgctgctg	551		
QY	307	ggacctacatcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc	366		
DB	552	ggcctctatagctccttcttcttcttcttcttcttcttcttcttcttcttcttcttctt	611		
QY	367	atggtgccaagttaccttctgctgctgctgctgctgctgctgctgctgctgctgctg	426		
DB	612	atctcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	671		
QY	427	ggcccaagtcgaattccaggtcttcaacatgacacaaagagatgctggaaca	486		
DB	672	ggcccgacagcccttgacacacttcatatgacaaatgga-----	708		
QY	487	gaagccatgagagctgagagagctgacgtgtgacatgacatgacatgacatgacatg	546		
DB	709	-cagccagagatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	767		
QY	547	atccagaatggtctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	606		
DB	768	ttccagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	827		
QY	607	ttcacacgggggttcaatgagacggcgccgacagatccatgattggtgctgacatgac	666		
DB	828	ctttgtccagagctatgacacacagctgacgtgctgacgtgctgctgacagctgacgt	887		
QY	667	atcttcgagctgacacatccctcctcctcctcctcctcctcctcctcctcctcctc	726		
DB	888	ggtgtttggccttcactgctgacgacgacgacgacgacgacgacgacgacgacgacg	947		
QY	727	gacatttgcacaaacctcccccacacacacacacacacacacacacacacacacacac	786		
DB	948	gagctgctgctgac	1007		
QY	787	ggtgctcctcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	846		
DB	1008	gggggtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	1067		
QY	847	cccatccctcacaagatgatgtygtgtygtgtygtgtygtgtygtgtygtgtygtg	906		
DB	1068	ccgattatcccggggacgtgctgacgcttcatcggggacacacacacacacacacacac	1127		
QY	907	atgcccacaaagatcacatgacagatcgtygtgtygtgtygtgtygtgtygtgtygtg	966		
DB	1128	cttaagacacagatttgaggtatgctggtggacacacacacacacacacacacacacac	1187		
QY	967	gtgtgctgctgtygttctacagatggaagacatgatagacacagccttccctagcatc	1026		
DB	1188	gtgggccccccac	1247		
QY	1027	gtgagctcagcatcacaacctggtcctgagccgagaccctggcacaagaacagagctgac	1086		
DB	1248	gttgggtttgctcatgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	1307		
QY	1087	tggtgctgcaagcagagatgatgctgctgctgctgctgctgctgctgctgctgctgct	1146		
DB	1308	gtggac	1367		
QY	1147	aaaattcatgcatctgtgctgctgctgctgctgctgctgctgctgctgctgctgctg	1206		

Db	1368	CAGTGCCTTCCCCGTCGATGTTGCTTATGTCTGTGGAGACCTCGTACAGCAAGAAAGCACCGGGGGC	1427
OY	1207	aaatccaggtgagcagccctgtgtgtctctgtgtgtgtatgataccatgctgtgtctgt	1266
Db	1428	AACTCGCGAGGTGTGCTGGAGCCATCTCTTCCCTTTTCATCTTCCTCATCTGTCATAACTT	1487
OY	1267	gggatactatctgataccctctccctcctgaagctgtgtcgaaggagccctgatctgttaatctc	1326
Db	1488	GGGGAACCTCTTCATGACACTGCCCAAGCGGTCTCTGGCAGCCATCATCTATTGTGAACTG	1547
OY	1327	aagaactccctcaagaacataccagaccctactaactcctgtgtgaggaagagcaagcttgac	1386
Db	1548	AAGGCGATGCTGAGGACGACCTAGCAGACATGCGCTCCCTCTGGAAAGCCAAATCGGCGGAT	1607
OY	1387	tggtgcatctgtgtgtatgactctctctccctctctctccctcctcctcagctcctatgtgtg	1446
Db	1608	CTGCTTATCTGGCTGGTGTACCTTCACGGCCACCATCTTGCTGTAACCTTGACCTTGGCTTG	1667
OY	1447	gcagtgaggtgtgtcgtcgtcctcgtccctgtgtgtgtgtgtgtcctcgaagactatctgaatgac	1506
Db	1668	GTGCTGGGCGTCATCTTCTCCCTGTGTCGCGTGGTGGTCCGAGACACATGCCCCACATAC	1727
OY	1507	tatgcatacgtcccaagtgatacgaactgtacattatgtgtgaatcccaagactataatg	1566
Db	1728	TCTGTCTTCTGGGGGACGCTGCACAGCAGATATATTATACAGAGATGTGGCAGAGTACTCAGAG	1787
OY	1567	gcccgagatataccagggatataaactatcaagtactgtcctccctctactcttccagac	1626
Db	1788	GCCAAAGGAAGTCCGGGGGTGAAGTCTTCCGCTCTCGGCTCACCGCTGTACTTTGCCAAT	1847
OY	1627	tcaagagatcttcag	1640
Db	1848	GCTGAGTTCTACAG	1861

RESULT 4

AF136751 LOCUS AF136751 2629 bp mRNA ROD 02-AUG-1999

DEFINITION Mus musculus down-regulated in adenoma protein (Dra), complete cds.

ACCESSION AF136751

VERSION AF136751.1 GI:5359729

KEYWORDS house mouse.

SOURCE Mus musculus

ORGANISM Eularivota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2629)

AUTHORS Melvin,J.E., Park,K., Richardson,L., Schultheis,P.J. and Shull,G.E.

TITLE Mouse down-regulated in adenoma (Dra) is an intestinal C1(-)/HCO(3)(-) exchanger and is up-regulated in colon of mice lacking the NHE3 Na(+)/H(+) exchanger

JOURNAL J. Biol. Chem. 274 (32), 22855-22861 (1999)

MEDLINE 99357827

REFERENCE 2 (bases 1 to 2629)

AUTHORS Melvin,J.E., Park,K., Richardson,L.A., Schultheis,P.J. and Shull,G.E.

TITLE Direct Submission

JOURNAL Submitted (23-MAR-1999) Center for Oral Biology, University of Rochester, 601 Elmwood Avenue, Rochester, NY 14642, USA

FEATURES

source

1..2629

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/sex="female"

/tissue_type="colon"

/dev_stage="adult"

1..2629

/gene="Dra"

131..2404

/gene="Dra"

[illegible]

[illegible]

RESULT	5				
AF230376	AF230376	4113 bp	mRNA	ROD	24-MAY-2000
LOCUS					
DEFINITION	Meriones unguiculatus prestin (Prest) mRNA, complete cds.				
ACCESSION	AF230376.1	GI:8050589			
VERSION					

KEYWORDS	REFERENCE	FEATURES
SOURCE	1 (bases 1 to 4113)	1. .4113
ORGANISM	Mongolian gerbil, Meriones unguiculatus	/organism="Meriones unguiculatus"
	Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Gerbillinae; Meriones.	/db_xref="taxon:10047"
	1 (bases 1 to 4113)	1. .4113
AUTHORS	Zheng, J., Shen, W., He, D. Z., Long, K. B., Madison, L. D. and Dallos, P.	
TITLE	Prestin is the motor protein of cochlear outer hair cells	
JOURNAL	Nature 405 (6783), 149-155 (2000)	
REFERENCE	2 (bases 1 to 4113)	
AUTHORS	Zheng, J., Shen, W., He, D. Z., Long, K. B., Madison, L. D. and Dallos, P.	
TITLE	Direct Submission	
JOURNAL	Submitted (01-FEB-2000) Endocrinology, Northwestern University, Tarry 15-703, 303 E. Chicago Ave., Chicago, IL 60611, USA	
FEATURES	Location/Qualifiers	
source	1. .4113	
gene	/db_xref="taxon:10047"	
CDS	1. .4113	
	/gene="Pres"	
	224..2458	
	/gene="Pres"	
	/codon_start=1	
	/product="Prestin"	
	/protein_id="AA071715.1"	
	/db_xref="GI:8050590"	
	/translation="MDAAENEIPVATOKYVERPIFSPHVLQERLHVKKVSDSID KLRQCTPKIKRINITYMPLPIRKMLPAYKREKVLADLVSISGLVLOSLAFR MLAAVPEYEGISFPYPIKCFPGTSRHSISGPNVISMIGVAVRLVDPDVIYIG GVANATNGEARDLAVKRVAMSVTLLSGIDCLVCRGFATILYELVKGFTTAA VHVETSMLEKLVGKTRKSGISFVSVYVAVALNKNVNSGLVLMFGLLGK EFNRFKPEKLPAPIPLEFFAVVMGTGSIAGNHLHESYSDVWGLPLGLPPAPDPDS LEHLYVDALIAIVGSEVITSMKTKNKGYSVDNDELIALGICNSIGSLPOTFS ISGSLISVQEGTGKTOAGLASMILLIVLATGEFLPESLPQVLSAIVLNLG MEMFSLPEPMPRTSKETLQITLTPFSLLFDGLDYLITAVIALLTVIRQSPSY KVLQQLPDTDVIIDIDAEYKEKLPKIKITDINAPITYANSDLKSKRTGNPML IMGARRKMRKYAKEVGANJANAIVAVKVDDEVGENATPREEEDVKYPPVTKTT FPEELQRMPTQENVHTIIDFTGVNFIIDSGVYTTLAVMKEXGDVGIYYVLACSPQ VAVNDLTNRRFFENPALKELLFHSIHDAVLGSHVREMAAEQASAPPDMEPNAPT TPFA"	
BASE COUNT	1053 a 995 c 1011 g 1054 t	
ORIGIN		
Query Match	15.8%; Score 358; DB 10; Length 4113;	
Best Local Similarity	51.9%; Pred. No. 1.3e-65;	
Matches	836; Conservative 0; Mismatches 765; Indels 9; Gaps 1;	
0Y	54 cctctcgacgatgatttgagaaagaagacccgagacataccagatggagagaactcgcg 113	
DB	304 CCTCCAGAGAGAGCGTCGATGTCGAAGACAAAGTCACAGTCCATTTGGGATTAACCTGAA 363	
0Y	114 caatgccttcgaatgttccctcagccaagaatcaaacgctgcyggttctggcctgcgcctcgt 173	
DB	364 GCAGCGCTTACATCTCACTCCCAAAAAGATAAGAAACATCTTTACATGTTCTCTCCCAT 423	
0Y	174 gctcctcgtgcctccccaagtaacaagaataaagaactacatacctccgacctgcgtcgtg 233	
DB	424 CACTAAGGCGTGTGCACAGCTTACAAAGTTCACAGAGATATGTTGGGTGACCTTGCTTCAGG 483	
0Y	234 actaagcggggatccatcccaaggttcccaagaagcattgcatcttgctgcctgcggccaact 293	
DB	484 CATAAGGACACGGCGTGTTCAGCTTCCCAAGGCTTACCTTCGCAATGCTCGCGGCTGT 543	
0Y	294 tccctgaatcaatggcctcactcctcctcttcccccctccgacctcactctccctgg 353	
DB	544 TCTCTCGGCTGTTGGCGCTGTACTCTTCAATTTTATCCGTATTATCAATGACTCTTTCTTGG 603	
0Y	354 ggggtgttcaacagatggtgcagtgacacttgcgtatccagcaacctgctggtgtaacat 413	
DB	604 GACCTCCAGACACATATTTATAGTGCTTTTCGCGCTCAATTAAGTTGATGATCGGGTGCT 663	

```

Oy 414 ctgtctcagcttgcccccagagctcgaaattccaggcttccaaatgcccacatlgagag 473
Db 664 GGCTGCTCCGGCTGCTGCCCGA-----TGACATGCTATCCCGGAGAGATGAACGC 714
Oy 474 ctatgttgacacagacagccatgagagctgtaagagctgacgtgtgtcagctagcctg 533
Db 715 AACCAACGGCAGGAGCCCGAGCGCCCTGAGATGAATTCGCATGTTGTGACCCCT 774
Oy 534 cctcaccgccatccatccagatgaggtctgagcttcatgcatgagcttggtgtgcccacta 593
Db 775 GCTCTCAGGAATCATTTAGATTGTTGCTTGTGCTGAGGTTTGATTTGGCCATATA 834
Oy 594 ccttcgcagagcttccatccggggcttcatgaagcgccggcctgacagatctgatttc 653
Db 835 CCTCAGGAGCGCGCTGGTGGAGGTTTCACACCGCCGCCCGGACGCTTTCACATC 894
Oy 654 ggtgtcagaatcatcttcgagctgacatccctccctacacagggccaggtccatcgt 713
Db 895 CATGTTGAAATACCTGTTGGGTTAAGACAAAGCGGTACGTTGGATCTTTGCTGCTG 954
Oy 714 cttaacctcatgacattgcaaaacctcccccacacacatcgctcgctcatcctt 773
Db 955 ATATAGTACACTGCTGCTGTGCGAATGTAAAACTCAACGTGTGCTTCCCTAGGCGT 1014
Oy 774 cgtctcatcagcgtgtgctctcgtgtgtgtgtaagagctcaatgctcgtacatgca 833
Db 1015 CGGCGCTGATGTTTGGTTTGGTGTGGTGGAGGAGGTTTATGAGAGATTTTAAAGA 1074
Oy 834 caagatcgccttccccatccctacacagagatgtgtgtgtgtgtgtgtgtgtgtgtgt 893
Db 1075 GAAATTTGCCAGCACCTCTCTAGAGTTTCTTGTGCTGTGTGTGTGTGTGTGTGTGTGT 1134
Oy 894 cgggggctgtaagatgtcccaaaaagtatcacatcagatcgctgagagaagaaaccaaagcg 953
Db 1135 CGCGGGGTTTAACTTGCACGAGTCTTACAGTGTGATGATGCTTGTGAACTTCTCTG 1194
Oy 954 gtcccaaccccggtgtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1013
Db 1195 GCTACTCCCTCCGTCGCAACCGGACACGACCTCTTCCACCTCGTGTGTGTGTGTGTGTGT 1254
Oy 1014 ctcctcagcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1073
Db 1255 TGCATGACCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1314
Oy 1074 gcaagcgtacagctgtgattcgaaacacagagatgagctgtgctgacgaacttctt 1133
Db 1315 GCATGGCTACCAAGTTGATGGCAATCAGGAGCTCATCGCTTGGGAGATATGCAACTCCAT 1374
Oy 1134 tgcctcctctttaaataatcgtcaattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1193
Db 1375 CGGATGCTCTTCCAGACCTTCTCATATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1434
Oy 1194 tggagctcggagagaataatccacaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1253
Db 1435 GGGGAACGTGAGGAGAAACACAGCTCGCAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1494
Oy 1254 catcgtgtccttgaggatcactcgtatccctccctcaagctgtgtcagagagcccgat 1313
Db 1495 CATTTTACCCACTGATTTCTCTTGAAGTCAATGCCCCAGGCTGTGCTCTGCGCCATTTGT 1554
Oy 1314 cgtgtcactatcaagaatccctcaagaacacacacacacacacacacacacacacacac 1373
Db 1555 GATGCTGAACCTGAAGAAGGATGTTTATGCACTGTCTGCAATGTGCCCCCTTCTTGAGAAC 1614
Oy 1374 gaggaaagctgacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1433
Db 1615 CAGAAATATAGAGCTGACCACTGCTGTATCCACCTTTGTGCTGCTGCTGCTGCTGCTGCT 1674
Oy 1434 gacctatggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1493
Db 1675 GGACTAGGAGCTGATGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1734

```

```

Oy 1494 gttcogaaatgctatgcatgctggccaggtcatgacactgacattatgtgaaatccaa 1553
Db 1735 GAGTCGCGATCAAGAGCTCTCCGGGACACTCCCTGACACCACTGATGATGACATAGA 1794
Oy 1554 gacctataataggcccccaggtatccagggatgtaaaatcaacgtactgtctccctt 1613
Db 1795 CGCATATAGAGAGGTGAAGAAATTCCTGGAATTAATAATATTCAGATTAACGCCCCCA 1854
Oy 1614 ctacttgcacactcagaagacttcaagcaaaaggtcatcgcccaagactg 1663
Db 1855 TTACTATGCAACAGTACTGTATAGCAACGCCCTTAATAAAGAAAGACTG 1904

RESULT 6
AY032863 2612 bp mRNA ROD 03-AUG-2001
LOCUS Mus musculus chloride-formate exchanger mRNA, complete cds.
DEFINITION AY032863
VERSION AY032863.1 GI:15080863
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 2612)
AUTHORS Knauf, F., Yang, C.L., Thomson, R.B., Mentone, S.A., Giebisch, G. and Aronson, P.S.
TITLE Identification of a chloride-formate exchanger expressed on the brush border membrane of renal proximal tubule cells
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (16), 9425-9430 (2001)
PUBMED 11459928
REFERENCE 2 (bases 1 to 2612)
AUTHORS Yang, C.-L. and Aronson, P.S.
TITLE Direct Submission
JOURNAL Submitted (20-APR-2001) Internal Medicine, Yale School of Medicine, 333 Cedar Street, P.O. Box 208029, New Haven, CT 06520-8029, USA
FEATURES
source
1. .2612
Location/Qualifiers
/organism="Mus musculus"
/strain="BALB/c"
/db_xref="taxon:10090"
145..2352
/note="CFEX; transporter; possible ortholog of Homo sapiens SLC26A6"
/codon_start=1
/product="chloride-formate exchanger"
/protein_id="AAK51131.1"
/db_xref="GI:15080864"
/translation="MGQQRDHYHVERPLNQEDLEGHMGPAKTHQWRTPFRCNSRA
RAHSLIQHVPVLMPLRPYPXREWLIDLSGLSVAIMQPLQGLAALLGLPMEGL
YSSFPVEITFELFETSRHSISGTFPAVSVAMGVSYSLESLTDKRAVQGLNATDARVO
VAYTLISGLFQVGLHVPVYTLISPELVYSYTTAASVOYLASQALVGVGKLS
SHSGPLSVITVLEVCAQDLPETVNGTYVTATVAVGVALVYLVKLRLEKRLPLPIPE
LTLTIGATGISYGVKLDNRKVDVGNITGLIPVAPKTELEFATLVGNMFAIAVGF
AIVASIGKIFALRHGVRVDSNOELVALGSLNIGFPCFPCPSVMSRSLVOESTGN
TOYAGVASSFLILIIYKLGELFEDLPKAVLAAYIYNLGMKMGKOFSDISLYKANRY
DLILWITFVATILINDIGLAVSIVSLIVVYRMOLPHYSVLGOVDPDIDYDVAE
YSGAKVEPGVAVPERSATLYFANAEISDSILKEKGVDBRLITQKKRKIKQEMKIK
RMKRAKSKODASAKTSISVSVYNTNLDPVKSNDVSGSAKYNQCELDIYVSSNE
DKAPITWLSKSLDLPQFHSILIDLSLSTISVDPVTCISIKNIIFRPREIEVEVYIA
ACYSPPVAQLEAGHFDEFSITKQHVFSVSDAVFALSHKRSVKSPLVATKI"
BASE COUNT 542 a 725 c 701 g 642 t 2 others
ORIGIN
Query Match 15.7% Score 355.8; DB 10; Length 2612;
Best Local Similarity 52.8% Pred. No. 4e-65;
Matches 808; Conservative 1; Mismatches 693; Indels 27; Gaps 1;
Oy 112 cgcacatgctcagatgctcctcagccaaagatcaagctgtgtgtgtgtgtgtgtgtgtgtgt 171
Db 250 CGAACTTGTTCCGATGCTCCCGCTCGGGCCCACTTCTTGTGCTCCACGACAGTTCCA 309

```

[illegible]

Db	Accession	Gene	Protein	Length	Score	E-value	Identity	Similarity
Db	1363	ATTTATCGTCAAACTTGGGGAACTCTTCCGAGACCTGCCCCAAGGCCGCTCTGCTCTGTC		1422				
Oy	1312	atcgctgcaatctcaagaactccctccaagaactcaacccactactactgtygaag		1371				
Db	1423	ATTATTGTGAACCTTAAAGGGCATGTATGAAGCAGTCTCTCAGACATCTGCTCTTTGGAAG		1482				
Oy	1372	aagagcaagctggaactgtgtgcatctcgggtaagagaactcctcctcctctcctcgc		1431				
Db	1483	GCAAACAAAGTGGACCTCTATCTGGGTGGTGAACCTTTGGCCACAAATCTGCTGAAC		1542				
Oy	1432	ctgcacctatgtgtggaagtygtgcacctctccctcctcctgctgctgctcagaact		1491				
Db	1543	CTTGACATTTGGGCTGTGGCAGTTTCCATAGCTCTCTCTCTGCTGCTGCTCCGAAATG		1602				
Oy	1492	caatttcgaatgcatgcatgcatgcccgaagtcatagcaactgacatlatgtgaatccc		1551				
Db	1603	CAGCTGCCCATTTACTCCGCTCCTCGGGGAGGTCGACAGATTATTAATAGAACGTG		1662				
Oy	1552	aagacctataatagggcccgagatctccaaggagataatcatcaagtlactgctccct		1611				
Db	1663	GCAGAAATACTCTGGGGGCAAGGAGGTCCCGGGTGTGAAGTCTTCGTCTCAAGCAGG		1722				
Oy	1612	ctctacttgcacactcagaagatctcag		1640				
Db	1723	CTGTACTTCGCCAATGCTGAGCTCTACAG		1751				
RESULT	7							
LOCUS	AF167412	2853 bp	mRNA	ROD	29-AUG-1999			
DEFINITION	Rattus norvegicus pendrin (Pds) mRNA, complete cds.							
ACCESSION	AF167412							
VERSION	AF167412.1	GI:5802228						
KEYWORDS								
SOURCE	Norway rat.							
ORGANISM	Rattus norvegicus							
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;							
	Rattus.							
REFERENCE	1 (bases 1 to 2853)							
AUTHORS	Everett,L.A., Morsli,H., Wu,D.K. and Green,E.D.							
TITLE	Expression pattern of the mouse ortholog of the Pendred's syndrome gene (Pds) suggests a key role for pendrin in the inner ear							
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9727-9732 (1999)							
REFERENCE	2 (bases 1 to 2853)							
AUTHORS	Everett,L.A. and Green,E.D.							
TITLE	Direct Submission							
JOURNAL	Submitted (09-JUL-1999) National Human Genome Research Institute, National Institutes of Health, 49 Convent Drive, MSC4431, Bethesda, MD 20892-4431, USA							
FEATURES	Location/Qualifiers							
source	1..2853							
	/organism="Rattus norvegicus"							
	/strain="Sprague-Dawley"							
	/db_xref="taxon:10116"							
	1..2853							
	/gene="Pds"							
	22..2364							
	/gene="Pds"							
	/note="Iodide/chloride transporter; Homo sapiens PDS is defective in Pendred syndrome (deafness and goitre)"							
	/codon_start=1							
	/product="pendrin"							
	/protein_id="AA021618.1"							
	/db_xref="GI:5802229"							
	/translation="MAARDRSEPPQALAEYSCSYAVSRPVYSELAQOQRERLRPERR							
	TLRDSLRSCSSGRKRAFKALPLPLDMLPKRYKWEMLSDIISGVSLGVLGGDGH							
	MAVALAAYVQYGLGAPFLLTLPVFGVSRHSVGPVPSLWQVLSNAPDDH							
	FLVPSGSLTINTTLDGTGRDARATLSTPLTLNLGIOLVGGIOLGIVRYLADP							
	LVGGFTTAAAFQVLYSOLKIVLAVSTRKNGVLSIYTLLEIFQNTGIDNIADFIAL							
	LTITVCAVVELNDPRFKHKLIPVPLPIEVITITLATAISGANLEAANNAGIVSIPSG							


```

/gorganism="Mus musculus"
/strain="BALB/c"
/db_xref="taxon:10090"
/chromosome="12"
/map="12B3"
1. .3086
/gene="Pds"
/gene="Pds"
217. 2559
/gene="Pds"
/note="Iodide/chloride transporter: Homo sapiens PDS is
defective in Pendred syndrome (deafness and goitre)"
/codon_start=1
/product="pendrin"
/protein_id="AA051617.1"
/db_xref="GI:5802227"
/translatiion="MAARGGSEPPQIAEYSCSYTSRPVYSSELAQQQRRRLPERR
TLRSLRSCSCSRKRAFGVKKLLPLIDMLPKRVKEMWLSDIISGVSTGLVTTLOGG
MAVYLAFAVPOETGVSAPFPIIYVEFGTSRIHSVPEPVSLMGSVSVLSMPDDH
FLVSGNGSALNSTLTLDGTGRDARVYLAFLTLVLTGIIQVLEGLDGFVRYLADP
LNGEFTAAAPFVYVLSOLKTLVNSTSTNGVLTSLIYTLFEIPONTGIDTNPADRLP
LTIIVCAVKEINDRFKRIIPVPIPIYITLTATISYCANLEKNTNAGTIVKSIPEG
FLPPLVPSGLGFSMDLASFSAIVAVAIASVGRVATVATHDVIDGNOETIAGISN
VFSEFSCFAMTALSTRPAVOESTGKTQVAGLISAIIVMAIVALGRLLLEPLOSVL
AAVYIAGKFMQVCDVPRLMKONKTDVAIWEFTCIIMSLIDGLLGLIFALLT
NVYLVQPSMNGLGSVPSTDIYKSIITHYKNLNEPEGVKILRFSSPIFYGNVDFKCI
NSYVGPDAIRVYNKRKLARKLRIOLIKKGRALRKNKILSIDSSNNAFEPDEVEEP
ELNLPITKEIIEIODYANSELDPVKYNVYKVPVHSLVLDGCAGVFLDYGVGNSLRATYVE
FORLDVAVFALLDDVLEKEWQCGFDDNDIKRDFLVTHADLLHQNQVKSREGDD
SLETVARIRCKCPDLDMEAKNAEELVDQDEAMRLAS"

```

Query Match	15.0%	Score 339.4	DB: 10	Length 3086
Best Local Similarity	51.7%	Pred. No. 1.2e-61		
Matches 803	Conservative	0	Mismatches 741	Indels 9
				Gaps 1
QY 120	cttcagatgttcctccagccaagatcaaaagcgtggtgttttggcgtgcgtgcgtcttc	179		
Db 375	CTGCAGTGGTCTCAAGAAAGAGAGCGCTTGGTGTGTGTTAAAGACTCTCTGCGCATTTCTGGA	434		
QY 180	cttggtccccaagtaacaagatlaaagactatacatcttcctgactcgtcgtgtgacac	239		
Db 435	CTGGCTCCCAAAATACCGAGTCAAGGATGGCTCTCAGTGCATCATCTCCGGAGTTAG	494		
QY 240	cggggagatccatccaggtcccaagaagatgcatcttgctctgtcgtggccaactcttcgc	299		
Db 495	CACGTGGCTGTGGGTACCTCTGCAGAGGATGGCTTATCCCTGCTGGGACAGATACGT	554		
QY 300	agtcattgctcttactc	359		
Db 555	TCAGTGGGTCTCTACTCTGTGCTTTCCTTTCCTTCCATCTGACGATATTTTGGTTCCGAAATC	614		
QY 360	tcaaccagatgtgtccaggtaccccttgcgttaccgaacacccctgtgtgtgtgtgtgt	419		
Db 615	AAGACACATCTTCAGTTGGCCCTTTCCTCCGCTGGTCAAGTTAATGGGGGATCGTGTGTCT	674		
QY 420	gcagcttgccccccagagtgtaaatctccaggtcttcaac-----aatgcaccaatga	470		
Db 675	GAGCATGGCTCCAGATGACCACTTTCCTTGTGCGCCAGCGGTAAACGGAATCTTGAATCTC	734		
QY 471	gagctatgtggaacagacagcatatgaggtctgaagagctgtgcaagctgtgacatagc	530		
Db 735	GACCAAGTTAGACATCTGGAACCAAGAGATCCGCGCGGATGTGCTTGGCAAGCACACTCAC	794		
QY 531	ctgcctcaacgacatccatccagatgtgtctgtggtcttcaagtttgacttgtgtgcat	590		
Db 795	TCTTCTAATTGGAATCATACAGCTGTTGTTGGAGAGTTTGACAGATTGGATTCAATAGTGAG	854		
QY 591	ctaaccttcgaagtccttccatcccggtgtctatgaagcgccgcccgcgtcgcagatctgat	650		
Db 855	GTACTTGGCAACACCTGTGGTGGGATGTCAACACCCCTCGACGCTTCCAAAGTACTGT	914		

DEFINITION	LOCUS	AY024359	2235 bp	mus musculus prestin mRNA, complete cds.	ROD	17-JUL-2001
Oy	651	ttgggtgctcaagaacatctctggagctgcagacatccctctcciaaagaagcccaaggtctat	710			
Db	915	ctcACAGCTAAAGTGTGTGCTCAATGTTTCAACCAAAAACCTCAACGCGATCTCTCCAT	974			
Oy	711	cgctcttaacctcaatctgacatttgcaaaaaccctcccccacacaacatcgctctgcatac	770			
Db	975	TATCTACACACTAATTATGATTTTTCAAATATTCGGTGACACCAATATTCGGGATTTCTAT	1034			
Oy	771	ctcgctctcaatcagcggctgcctccctcgtgctcgtgtgaagagatcaatgctcgtctaat	830			
Db	1035	CGGTGGGCTGCTGCATCATCTGCTGTATGAGCTGTAAAGAACTAAATGATCGATTAA	1094			
Oy	831	gcaacaagattgcctccccctccctacagaagatgattggtgtgtgtgtgtgtgtgtgtgtgt	890			
Db	1095	ACACAGAAATCCCGGTGGCCATTTCTTATAGAAGTGATTTGTGACAAATATATGTATCGCAT	1155			
Oy	891	ctccgagggcgtgaagatgcccacaaaagtatcacatgcagatcgctgtgtgtgtgtgtgtgtgt	950			
Db	1155	TTCTCTATGAGGCCCAACTTGGAAAAAGAACTCAATGCTGGCATTTGTTAACTCCATCCAA	1214			
Oy	951	cgaggctcccccaccccggtgtgtcgcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1010			
Db	1215	TGGGTTCTTGCTCTCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1274			
Oy	1011	ctctccctcagacacgtgtgagctacgtatcaacacctggtctatgtgtgtgtgtgtgtgtgtgt	1070			
Db	1275	CTTTTTCATTTGCTGT	1334			
Oy	1071	caagcacggcgtgaagcgtgtgattcgaacacagaagatgatacgtctcgtgtgtgtgtgtgtgt	1130			
Db	1335	CAGAGTACTATATGTATCATCATGTGGAAACAGAGAAATTCATTGCTTTGGGATTAAGCAAGT	1394			
Oy	1131	ctttgtgcctctctcttcaaaatlcatatttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1190			
Db	1395	CTTCTCTGAGATTTTTTCGCTGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1454			
Oy	1191	ggatgtgagctgtgaggaataatcccaagttgtgcacgcctgtgtgtgtgtgtgtgtgtgtgtgt	1250			
Db	1455	GGAGAGCACCGGGGGGAAAGACACAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1514			
Oy	1251	caccatgctgt	1310			
Db	1515	TGCCATCGTTTCCCTTGTGGGAGGCTTCTGGAAACCTTGCAGAAATCAGTCTTGTGGGCGCT	1574			
Oy	1311	gacgcctgtcatatcacaagaatccctccaagaacacacacacacacacacacacacacacacac	1370			
Db	1575	TGTCTATTGCCAACTGAAAGGAAATGTTTATGACAGGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1634			
Oy	1371	gaagagcaagctgtgactgtgtcatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1430			
Db	1635	GCAGATATTAAGCATGATCTGTATATGTGGGCTTGTATCATGATTAATGTCCATCATTTCTGG	1694			
Oy	1431	ctgtgcccatagt	1490			
Db	1695	GCTGGACCTGGGCTTGT	1754			
Oy	1491	tcaattctcgaatgtgcatgtgacatgt	1550			
Db	1755	TCAAGTCCCTTCAATGAGAAATGGCCTTGGAAATGTGTCCACACAGCAATCTACAAAAGCAT	1814			
Oy	1551	caagacatataataggt	1610			
Db	1815	CACACATATTAATAAACTTGAAGAGCTGAAGGGGTGAAGATCTTCAGATTTTCCAGTTC	1874			
Oy	1611	tctctactcttgcacacccaagaatctctcaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1663			
Db	1875	CATTTTTTCAGCGCAATGTCAATGCTTTTAAAAAATGTATCATATTCAACAGCGTTG	1927			
RESULT	9					

ACCESSION AY024359
 VERSION AY024359.2 GI:14787222
 KEYWORDS
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 2235)
 AUTHORS Adler, H.J., Merritt, R.C. Jr., Belyantseva, I.A. and Kachar, B.
 TITLE Mus musculus organ of Corti mRNA for prestin
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2235)
 AUTHORS Adler, H.J., Merritt, R.C. Jr., Belyantseva, I.A. and Kachar, B.
 TITLE Direct Submission
 JOURNAL Submitted (25-JAN-2001) NIDCD, NIH, Bldg. 36/5D15, 36 Convent
 Drive, Bethesda, MD 20892-4163, USA
 COMMENT On Jul 17, 2001 this sequence version replaced gi:12658642.
 FEATURES
 location/qualifiers
 1..2235
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /tissue_type="organ of Corti"
 1..2235
 /codon_start=1
 /product="Prestin"
 /protein_id="AAG59999.2"
 /db_xref="GI:14787223"
 /translation="MDHAENEIPAEIORYVERPISHPVLOERLHVKKVTSIGD
 KLAQFTCPKIKNIITIMPLPIKMLPAKPKRYVLDLVSIGSTGYLOLPQGLAFA
 MIAVPPYFEGISSTPYVIMKCFSTSHISIGPAVLSIMIGAAVLPDDIYIG
 GVNAFTNGEADALRVKAVSVTLTSGIIOCLGCRGFAIYLEPLVGFPTAA
 VHVTSMLKLYGVKTRKYSIGFISVYSTAVLVNKNLNCISLGVLMVGLLGK
 EFNRFEXKLPAPLEFPAVMTGTSAGNLEHSYVDVYDPLILPANDTS
 LEHLYVDALIAIVGESVTISMAKTLANKGYODGNOELIAGICNSISLPOTS
 ICSLSRSIOEGTGKQIAGCLASMLIIVITATGPLESPISPAVLSAIVYLNK
 MEMQSPDPFTWRTSKIEPTIMUTTPSSLEGLDYGITAVIALLVYTRTOSPSY
 KVLQGLPDTDYIDIDATEVKEIPGKIFQINAVRYANSDLSSALKRGTVPAL
 IMGARARMKRYAKEVGNANVANAIVKVAEDVGENATKPEEDDEVKFPYVITKT
 FPELQRLPGQENVHVIILDTQVNEFVDSGVGTLAGIVKEGVGIVYVLAGSPQ
 VVNDLFRNFPENPAKLELPHSHIDVAISGVNEMAMEQENATSLPQEDMEPNATPT
 TPEA"
 BASE COUNT 587 a 516 c 542 g 590 t
 ORIGIN
 Query Match 14.9%; Score 337.2; DB 10; Length 2235;
 Best Local Similarity 51.7%; Pred. No. 3.5e-61;
 Matches 821; Conservative 0; Mismatches 758; Indels 9; Gaps 2;

436 tcgaatccaggtcttcaacaatgcacccaatgaagactatgtgacacagaacatg 495
 463 GATA-----TTGTCATCCAGAGAGAGTAATGCAACCAAGGG---ACAGAGCCAGA 513
 496 gaagctgaagagctgaagctgacgtacgtacgtacgtacgtacgtacgtacgt 555
 514 GATGACCTAAGAGTGAAGTCCCATGTCGTACCTTACCTTACCTTACCTTACCTT 573
 556 ggtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 615
 574 TCCCTAGGCTGTAGCTTGTAGCTTGTAGCTTGTAGCTTGTAGCTTGTAGCTTGT 633
 616 ggtcctacagc 675
 634 GCGTTTACCACTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 693
 676 ctgaccatccctcctacacagagccagaggtccatcgtcttacccttaccatgttc 735
 694 GTCAAAACAAACGCGTACAGTGAATCTTTACAGTGTGTATAGTACACTTCTGTG 753
 736 aaaaaccccccaac 795
 754 CAGAAATGTTAAACCTCAACCTGATGCTTCCCTAGCGCTGCGCTGATGCTTTGGTTG 813
 796 ctggt 855
 814 CTTGTTGGGTGACAGAAATTAATAGAGATTTAAAGAGAAATTTGACACACCACTTCT 873
 856 aagaagatgattgt 915
 874 CTAGATCTTTGCTGT 933
 916 aagttacatgcagatcgt 975
 934 TCTTACAGT 993
 976 gtgtgtcagatgt 1035
 994 GACACACAGCTTCTCACCCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1053
 1036 gtccatcaactgt 1095
 1054 TCAGTACGATCTTCATGCGCAAAACCTTGCAAAATACCATGCTACAGGTGTAGTGC 1113
 1096 aaccagagatgatcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1155
 1114 AATCAGAGACTTATGCTTGTGGGATATGCAACTCCATTGGATCTCTTCCAAACCTTC 1173
 1156 gtccattgt 1215
 1174 TCGATTTCCTGCTCCTTGTCTGGAACCTTTGTCAGAGAAAGAACTGAGGAAACACAG 1233
 1216 gtgcagcagctgt 1275
 1234 CTTCGAGGTGTTGGCCCTGTTGATGATCTTGTGCTATTAATCCACCGGATCTC 1293
 1276 ctgtatccctccctaaagctgtgtcagagccctgaatgcgtcgtcgtcgtcgtcgtc 1335
 1294 TTTGAGTGTATACCCAGGCTGTCTTTCGCCCATTTGTATGTCTCACTGAAAGAAATG 1353
 1336 ctcaagcaactacacgacccctactactcgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1395
 1354 TTCATGACATCTCAAGACTGCTGCTTTTCTGAGAACACGAAATATAGAGCCGACATC 1413
 1396 tgggt 1455
 1414 TGGCTGACACCTTGT 1473
 1456 gtgcgtcctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1515
 1474 GTGATCATTTGCTCTGCTACAGTGTATTTATAGAACACAGAGTCAAGCTACAAAGTCT 1533

```

Oy 1516 gccacggtcatgacacatgacattatgtgaatcccaagaccataataggcccaagat 1575
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1534 GGGCAGCTCCCTGACACGAGTGTGTACATTTGACATGATGACATGAGGAGTGAAGAA 1593
Oy 1576 atccagggtatataatatacgaactctctccctctactcttgcacactagaatc 1635
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1594 ATTCTGGAAATTAATAATATTCCTCAATTAATGCCCAATTACTATGCAATAATAGCGACTTG 1653
Oy 1636 ttcaggcaaaagtcacgcacaaagctg 1663
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1654 TATAGCAGCGCTTTAAAAAGAAAGACTG 1681

RESULT 10
AF314819 1389 bp mRNA MAM 12-FEB-2001
LOCUS AF314819
DEFINITION Oryctolagus cuniculus down-regulated in adenoma DRA (SLC26a3) mRNA,
partial cds.
ACCESSION AF314819 GI:12656493
VERSION AF314819
KEYWORDS
SOURCE
ORGANISM
Oryctolagus cuniculus
    rabbit.
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
REFERENCE
    1 (bases 1 to 1389)
    Jacob, P., Rossmann, H., Lamprecht, G., Kretz, A., Neff, C., Lin-Wu, E.,
    Gregor, M. and Seidler, U.
    Characterization of rat and rabbit brush border membrane anion
    exchange and DRA expression in rabbit, rat, and human duodenum
    Unpublished
    2 (bases 1 to 1389)
    Rossmann, H., Jacob, P. and Seidler, U.
    Direct Submission
    Submitted (16-OCT-2000) Abteiling Innere Medizin I, University of
    Tuebingen, Oefried-Wueller-Str. 10, Tuebingen 72076, Germany
FEATURES
    source
        1..1389
        /organism="Oryctolagus cuniculus"
        /db_xref="taxon:9986"
        /tissue_type="intestinal mucosa"
        <1..>1389
        /gene="SLC26a3"
        <1..>1389
        /gene="SLC26a3"
        /note="anion exchange protein; member of the SLC26 gene
        family"
        /codon_start=1
        /product="down-regulated in adenoma DRA"
        /protein_id="RAK00897.1"
        /db_xref="GI:12656494"
        /translation="EPISWLPREYKLEMLSDIVSGISTGLVAVIAGLAFALVNIP
        PAYGLAYAFPIYVFFPGTSKHISVGPVLSMWVAVVLAADDISIPSPNSNT
        DDLNLRINVAASVTEFLITGLAFLIRIGFVIVISPAISLIFGTAALAHVISO
        LKPMOLTPVPAHTDPSIFKVLDESFTQIRKTNADLYTKRIIVIVYPRKEINER
        KAKLPVPPIELIVIVIAAGSYGCFQQRNVSIVIGMEGFOAPAPDQVEDAI
        GDCFTIAIVGQVAVASVSVSLKYDIRIDSQELIAGLQIVTSGSKGSGSTALS
        RSAVGESTGKTQVAGVLSVILVIAFLLEIVLFLQKSVLSALALNKLMLQFA
        EIGRLMKDKDKDCLIMTFTFAIVIGLGLASVARELLIYVRQEPKCTTLANI
        GRSNIYKRKRKXYS"
BASE COUNT 319 a 363 c 345 g 362 t
ORIGIN

```

```

Query Match 14.8% Score 334.6; DB 4; Length 1389;
Best Local Similarity 54.2%; Pred. No. 1.3e-60;
Matches 751; Conservative 0; Mismatches 619; Indels 15; Gaps 3;

```

```

Db 73 AGCAGAGCGGTGAGTACTGCTGCTCAAGGTTTAGCTTGTGACTGCTGTCACATCTCCC 132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 298 gcaatcaatgagctctctacccctctctctccctctcgaactactctctccgggggt 357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133 CCAGCTACGGGGCTATATACAGATATTTCCCGGTATAGTATCTTTCTTTGGCAGCT 192
Oy 358 gtccaccagatggtgcgaagtaactctgcgcgtatcaagatccctgggtgaacatctgt 417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 193 TCTAAACACATCTCTGTGGTCCATTTCGGGTTCTGACATGATGTTGGGAGACAGATT 252
Oy 418 ctgcagctggccccagagtcgaaatccaggtctcaacaaatgcacaaatgagagctat 477
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 253 GTGAGATTAGCCCCAGAG-----CGACAGTATATCCACAGACCCCTTAACAACACTCG 303
Oy 478 gtggaaacagcaagccatggagagctggaagctgcgaagctgcatgctcgtacgtccctc 537
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 304 ACAGA--TGATTCACTGACAAACAGAGAGATATAGTGGCTGCATCACTCACTTCCTC 360
Oy 538 accgcatcatccagatgggtctggtctcaatgacgttgctgtgtgccaatccatcc 597
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 ACTGGAATCATCAAGCTGGCTTTGGCATTTCTCCGATTGGGTCTGGTGAATCTACTG 420
Oy 598 tccgagctctcatccggggtctcatgacgcccgcgcgtgcagatcctgattctcggtg 657
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 TCGAGAGCCCTGATACAGCCCTTACACACTGCGCGCGCACTCCACTGTGTGATCTCCAA 480
Oy 658 cttaagtaactctgagtagaccatccctccctcaagagcccaaggccatcgtcctt 717
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 481 CTCAAATTCATGCTGCAGCTGACAGTCCGGCACACACAGACCCGTTTCCATTTTCAAA 540
Oy 718 acctcaatgacaattgcaaaaaccccccacacacaaatcgcctcgtcatcctcgt 777
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 GTCCTGAGTCCCTATTTCACACAAATCCGGAAGACCAACATTCACACCTGGTGACAGG 600
Oy 778 c---tcaatgaaggtgctctctcgtgctcgtggtggaagagctcaatgctcgtcatgca 834
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 601 CGGCTATCATGCTGCTGATCGTGTGTTGTTTAAAGAAATAAACAGAGCGCTTCAAGGC 660
Oy 835 aagatcgtctcccaatccctcaagacatgattgtgtgtgtggaacacatcatcc 894
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 661 AAGCTTCCCGTGGCCATCCCATTTGAGCTCATTTGTATACCGTATTCACAGGCTTGTCC 720
Oy 895 gggggtctgaatgaccccaaaaagatcaatcatgacatcgtgtggagaatccaaacgcggg 954
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 721 TAGCGCTGTAACTTCCACAGCATTCACAGTTCAGTGTGATCGGGAATGAGAAAGGG 780
Oy 955 tccccaccgggtgtgcgctgtgtgtcctcaacgttgaagacatgataagcaagcttc 1014
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 781 TTCCAAAGCCCTGCTGCGCTCTACACACAGGTTTTCAGAGATGCCATTGAGAGCTGCTTC 840
Oy 1015 tccctagcatcgtgagctcaagctcatcaacatcgtgatagtgacggcccttgcacaaag 1074
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 841 ACCATTGCGATGTGGGCTTTCGTCGTCGCTTTCCGTGACAGCCTCATTTCCCTCAAA 900
Oy 1075 caaggtcagcagctggtatcgaaacccagagatgcatcgtcctcgtcgcagcaactctt 1134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 901 TAGCACTATTCGATTGACCGCAGTACAGATTAAATAGCTTGGCCTGGGTGAACATCGTC 960
Oy 1135 ggcctcctcttaaaatcaatgcatctgtcgtcgtcttctgtaactctgtaactgtgtg 1194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 961 ACTGATCGTTCAAAAGGTTTCCCGAAGTACCGCCCTTCCCGCTTCGGGTTCAGAGAG 1020
Oy 1195 ggaagctgagaagaatcccaagctggaagcctgtgtgtgtgtgtgtgtgtgtgtgtgt 1254
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1021 AGCAGAGAGGAGGAAAACGAGAGTTGCGGGGTTCTGTCTTCTGTATCTGCTAATTGTC 1080
Oy 1255 atgctgtgtcctgggagatcatctatcctcctccaaagctcgtgtgaagagccctgatac 1314
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1081 ATTGTAGATATTGGAATTCTTCGAGGACCATACAGAAATGTCTGCTGTCACAGCTTAGCA 1140
Oy 1315 gctgtcaatctcaagaactccctcaagcaatccacgcacccactactcctgtgtgaggag 1374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db	1141	CTTGGAACCTTAAAGGGAATGCTGATGCAAGTTTGGCTGAGATAGGACAGACTGTGGAAAGAG	1200
Qy	1375	agcaagctgagctctgtgcatctgag99tagtgagctcctcctcctcctcctcctcctcagctg	1434
Db	1201	GACAAATACGACTGTTTAATTGATGATGACACTTCATCTTTCATCTTGGCAATGCTCTGGAGACTA	1260
Qy	1435	ccctatggtgtgagcagtgagtggtgtgcctctcctcctcctcctcctcctcctcagcagc	1494
Db	1261	GGATTTGGCTCTGGCGGCGCAAGGTGTGGCTGTTTAAACCTCCACCATCTGATTTAGAGACTCAG	1320
Qy	1495	ttctgcaatgctctgagcagctgcccagctcagcagcagcagcattgagtatcccaag	1554
Db	1321	TTTCCAAATGACACCACTGCTGACTAACATCGGAGAGGACACACATCTAATAGATGAGAAA	1380
Qy	1555	accta	1559
Db	1381	GACTA	1385
RESULT	11		
RNO303372			
LOCUS	RNO303372	2235 bp	RNA
DEFINITION	Rattus norvegicus mRNA for prestin.		ROD
ACCESSION	AJ303372		29-MAR-2001
VERSION	AJ303372.1	GI:12188917	
KEYWORDS	prestin.		
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;		
	Rattus.		
REFERENCE	1 (bases 1 to 2235)		
AUTHORS	Ludwig, J., Oliver, D., Frank, G., Klocker, N., Gummer, A.W. and Fakler, B.		
TITLE	Reciprocal electromechanical properties of rat prestin: The motor molecule from rat outer hair cells		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	98 (7),	4178-4183 (2001)
MEDLINE	21173705		
REFERENCE	2 (bases 1 to 2235)		
AUTHORS	Ludwig, J.		
TITLE	Direct Submission		
JOURNAL	Submitted (08-JAN-2001)		
	Ludwig J., Physiologie II / MNF, Universitaet Tuebingen, Ob dem Himmelreich 7, Tuebingen D-72074, Germany		
FEATURES			
source	Location/Qualifiers		
	1..2235		
	/organism="Rattus norvegicus"		
	/db_xref="taxon:10116"		
	1..2235		
	/codon_start=1		
	/product="prestin"		
	/protein_id="CAC21555.1"		
	/db_xref="GI:12188918"		
	/translation="MDHAEEKEIPAEFTOKYVERPFPSPVLOERLHVKKYDSIGEF		
	KLAAPFTCPKRVNIIYMLPIPKMLPAYFKKEYLADIGVLSGVLOPLGLAFRFGV		
	MLAAPVYGVSEFPYIMYICFPTGMSILGEPVAVSLGVAVALVEDDIIVIG		
	GVNNTNGEARDALRVKAVMSVTLISGLIOCLGCRGEFAVILTEPLVGGFTAAAR		
	VHVTSMILKTVFGVTKTKYSGIFSVVSTGVILNVKNLCISGVGLMVGILLGGK		
	ENRNRKELPAPLPLEFPAYVMGTGISAGNLEHSYVDVGTLPGLGLPPANDPDS		
	LFHLVYDAIAIATGFSVTISMKATLANRKGIVDGMQELIAGICNSIGSLCVTFS		
	LSCASRLVDEGTGKQTLAGCLASMLILVILATGTLFESLPQAVLSAIVYLNKG		
	MFMOFLDPFMWTSKIELITLMTTPVSLGIDYGLTAVIALLTVIRTPSSYSL		
	TVLODLPDPTVDIDIDAAVEEKEIPIGIFQINAPYIYANSDLAKRKGTVPAPAI		
	IMGARRKMRKRYAEKAVGNANANANATAVVDAEVDGENATREEDDDVKKPPYIKTT		
	FPEELQRLPDGGENIHYYIIDTPQVNRFDSDVGVTTIAGIVKEYGCVGIVYVLAGCSAQ		
	VVNDLTSNRFFENPALKEHLFHSIHDAVLSGVDRVAMAEQETVLPQEDMEPNATPT		
	TPEA"		
BASE COUNT	560 a	538 c	563 g
ORIGIN	574 t		
Query Match	14.7%	Score 332.8;	DB 10; Length 2235;
Best Local Similarity	51.5%;	Pred. NO. 3e-60;	

Matches	819:	Conservative	0:	Mismatches	762:	Indels	9:	Gaps	2:
QY	76	aagaagagaccggacac	taccacag	tggagagaaact	tcgcaatgcgcttcacagatgcttc	ca	135		
Db	103	AAGGACCAAAAGTCAC	AGACTCCATCGGGGAT	TAAGCTGAACGAGGACATTCACGTCAC	TCTCT		162		
QY	136	gccaaagatcaaaagc	tgtgtgtgtt	tgggcgtgcgtcgtgtc	tctcctcgtgtcctcccaagtac		195		
Db	163	AAAAAAGTAAGAAAC	ATCATCTACATGTGTTCTTGCCCATC	ACTAAAGTGCTTGCCAGCATAT			222		
QY	196	aagatlaaagatcatca	ctccttcgaaccgtgcgcgtg	tgagactgaagcgaggatccatccag			255		
Db	223	AAATTTCAAGGATAT	GTGCTGGGTGCTTGCTGTCCGAGAT	TAAGCACTGGGCTCTTCA			282		
QY	256	gtccccaagaagaa	lvgcattgtctgtcgtgcgcaaac	cttcctgtgcagtlcaatgtgcctac			315		
Db	283	CTCCCCCAAGGCTT	ATGCTTACCGATGCTGGCAGCTGTGCT	CTCCGGTGTTCGGCTGTAC			342		
QY	316	tctctcctctccccc	ctcctgaacctactctctc	ctgtgggggtgtgttcaagaatgtgtcca			375		
Db	343	TCTCATATTTTACCC	CGGTATTCATGTACTCTTTCTTTGGAACCT	CCAGACACATATCTATA			402		
QY	376	ggtactctgtccgtla	taacgalcctcgtgtgtgttaacalc	ctgtctgtgagctgcgcccaag			435		
Db	403	GGTCTCTTTTGGCG	TATTATAGCTTGATGTGGAAGTGCTGTGCT	GTCCGTGGGTAC-----			457		
QY	436	tcgaaatlccaaagt	ctctacaaatgcccacaa	lbgagagctatgttgagacaagagccatg			495		
Db	458	-GGGATGACATGT	TATCCCGAGGAGATTAATGCAACCAATGGG--	-ACGGAAGCCAGA			513		
QY	496	gaagctgaaaggtc	gaegtgtcagctacgtctgacctgc	ccacgcgcacatccagatg			555		
Db	514	GATGGGTTAAGG	GTCAAACTCCGCATGTCCGTTACCCCTT	TCAGAAATCATTCAGTTT			573		
QY	556	ggtcgtgggtatgca	ggtttgtgtgtgtgagcatctactc	ctccgagatccatccacgg			615		
Db	574	TGCTTAGGTTGCT	GTAGGTTTGATTTGGCCATATTA	CTCAGGAGCCGCTAGTGCCA			633		
QY	616	ggcttcatacgcgc	gcgcgcgtgcgaatcctgpatctc	gtgtgtcgtcgaagtacatcttcgga			675		
Db	634	GGCTTTACACG	CGCGCGGCTGTGCACGTTTCACCTC	ATGTTAAATATACCTGTTTGGG			693		
QY	676	ctgaccatccctc	ctacacaaagcccaaggtccalc	ctgtctcttaacttcattgacattgc			735		
Db	694	GTCAAAACTAAG	ACGGATACGTGGGATCTTCTCAGTGTGAT	ATGTAACCGTGCTGTGCTG			753		
QY	736	aaaaac	ctcccccaacacacacalc	gtcgtctgtcatcttctgtctctaaagcggtgcctc			795		
Db	754	CAGAAATGTTAA	AAAACTCTACACGTGTGTTCCCTTAA	GGCGTGGCCCTCAATGCTTTTGTGTTTG			813		
QY	796	ctgtgtctgtgtga	agagagctcaatgtcgtctacatgc	acaaagaltcgtctcccaaccct			855		
Db	814	CTGTTGGTGGCA	AGGAGTTTATGAGATTTTAAAGAGAA	ACTGGCAGGCCCATTTCT			873		
QY	856	acagagatgatctgt	gtgtgtgtgtgtgcacaaagctatc	ctccggygtcgtlaagatgtccaaa			915		
Db	874	CTAGAGTTCTTGT	GTGTGTGATGGGAGCGGCACTTTCTG	CAGAGGTTTAACTTGACAGAG			933		
QY	916	aagtatcaatcaga	tcgtgtggagagaatcccaacgcg	gggtgtcccaaccocgggtgtgcct			975		
Db	934	TCTTACAGGTG	AGATCTCTTTGGGACACTTCTCTTG	GGACTACTCCACCCGGCAACCG			993		
QY	976	gtgtgtctcacgt	tgaaagacatgatcaggaacgcct	cttcccttagcaatcgtgtgactac			1035		
Db	994	GACACACAGCT	GTTCACCTCGTGTACGTGACGTGAC	GCATTCGCCATTTGTTGGGTTT			1053		
QY	1036	gtcatcaacctgt	catctgtggccggaacctgtgcacaa	gcgaacgcgtctacgaactgtgactcg			1095		
Db	1054	TCAGTGAACAT	GTCCATGCGCAAAACCTTGCGCAAA	TAATGAAGCATGGTACACGAGTTGATGCG			1113		
QY	1096	aaccagagatgat	cgctctcgtgtcgtgcaggaactct	tgttgccctcttctaataatcat			1155		
Db	1114	AACCAAGACCT	CTATTCCTTGGGAATATGCAAC	TGCATGCAGATCTCTCTTCAGACCTTC			1173		


```

Db 144 --ALNDSMINETARDAAARVOASTLSVLVGLFQVGLGILHGFVVTYIASEPLVRGYTTAA 201
Qy 212 GLQILISLVKTYIFGLTIPSYTPGSIYFTFIDICKNLPHTNIASLIFALISGAVLVYKE 271
Db 202 AVQFVFSQKTVFGLHSHSGPLSLITVLEVCMKLPQSKVGVVTAAGVAVLVVYKL 261
Qy 272 LNAIRMKIRPIPTTEMIVVAVATAISGCGCKMPKKYHMOIVGEIQRGFPPTVSPVSGWK 331
Db 262 LNDLQOOLPMPPIGELTLLTGATGISYGMGLKHFREVDVGNIPAGLVPPVAPNTQLFS 321
Qy 332 DMIGTAFSLAVSYIYNAMGRTLANKRGYDVDSNOEMIALGCSNFGSPFKIHVICAL 391
Db 322 KLVSAFTIAYVGFALIAISLCKIFALRHGRVDSNOELVALGLSLIGITQCFPVSCSM 381
Qy 392 SVTLAVDAGAGKSOVASLCSVLVYMTLVGLIYLPPLKSVGLALLAVNLKSLQOLD 451
Db 382 SRSIVQSTGNSQVAGAISSLFILLITVLEVCMKLPQSKVGVVTAAGVAVLVVYKL 441
Qy 452 PYLMRKSKLDCCITWVVSFLSFLSLPYGAVGVAVSVLVVVFOTQFRNGTALAQVMDT 511
Db 442 MRSIMKANRADLLIMLVFTATILNLDGLVAVIFSLLVVVTQMPHYSVLGQVPT 501
Qy 512 DIYVNPYTNRAODIOGKITTYCSPLYFANSEIF-----ROKVIK 553
Db 502 DIYDVAVETSAKEVKGKVRSSATYVFANAEPYSDALKRCGVADVDFLSQKKKLK 561
Qy 554 --TVSLOELOD---FENAPPTDP-----NNNQTPANGTSVSYTFSPDS 593
Db 562 QEOLTKLOLQKEEKLKROAASPKGASVINVTSLSDMRSNVVEDCKMKQVS---SGDK 617
Qy 594 SSPAOSPEPAEAEPGESDMLASVPPVTFHTLILDMGVSFVDLMGIRKALAKLSIYTG 653
Db 618 MEDTANGQEDSKAP-DGSTLKALGLPPDPDFHSLILDGALSFVDYVCLSKLNFHDFR 676
Qy 654 KIGKVFVFNINIAQVYNDISHGVFEDGSLCKHVPFSDIHDAVYFAQANADV 706
Db 677 EIEVEVYMAACHSPVVSQLEAGHFF-DASITKKHLFASVHDAVTFALQHPRPV 728

```

```

RESULT 2
Q9BXK9 PRELIMINARY; PRT; 759 AA.
ID O9BXK9;
AC O9BXK9;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SOLUTE CARRIER FAMILY 26 MEMBER 6.
GN SLC26A6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21145584; PubMed=11247665;
RA Waldeger S., Moschen I., Ramirez A., Smith R.J., Ayadi H., Lang F.,
RA Kubisch C.;
RT Cloning and Characterization of SLC26A6, a Novel Member of the Solute
RT Carrier 26 Gene Family.
RL Genomics 72:43-50(2001).
DR EMBL: AF288410; AAK19153.1;
DR SEQUENCE 759 AA; 82966 MW; 63CB0B756C9675C6 CRC64;

```

Query Match 32.5%; Score 1258; DB 4; Length 759;
 Best Local Similarity 36.6%; Pred. No. 1e-70;
 Matches 261; Conservative 162; Mismatches 238; Indels 52; Gaps 8;

```

Qy 32 PVGEKLNARFCSAKTAVFGLPVLVSWLPKIKDYIIPDLGLSGSIOVPGMA 91
Db 51 PRHQMWTWIGCSARAYALLQHLPLVLMFRYPVBDMLGLDLSGSAVIMLPGIA 110
Qy 92 FALLANIPAVNGLYSSFFPLITFFLGGVHOMVPGTAVVISILVNICLOLAPESKQV 151

```

```

Db 111 YALLAGLPVVEGLYSSFFVPIYFLFTSRHISVGTFAVMWSMGVSTESLAPQ----- 164
Qy 152 NNATNESVYDPAAMAEELHVSATLACTAIITOMGLGMQGFVAIYSEFIRGEMTAA 211
Db 165 --ALNDSMINETARDAAARVOASTLSVLVGLFQVGLGILHGFVVTYIASEPLVRGYTTAA 222
Qy 212 GLQILISLVKTYIFGLTIPSYTPGSIYFTFIDICKNLPHTNIASLIFALISGAVLVYKE 271
Db 202 AVQFVFSQKTVFGLHSHSGPLSLITVLEVCMKLPQSKVGVVTAAGVAVLVVYKL 261
Qy 272 LNAIRMKIRPIPTTEMIVVAVATAISGCGCKMPKKYHMOIVGEIQRGFPPTVSPVSGWK 331
Db 262 LNDLQOOLPMPPIGELTLLTGATGISYGMGLKHFREVDVGNIPAGLVPPVAPNTQLFS 321
Qy 332 DMIGTAFSLAVSYIYNAMGRTLANKRGYDVDSNOEMIALGCSNFGSPFKIHVICAL 391
Db 322 KLVSAFTIAYVGFALIAISLCKIFALRHGRVDSNOELVALGLSLIGITQCFPVSCSM 381
Qy 392 SVTLAVDAGAGKSOVASLCSVLVYMTLVGLIYLPPLKSVGLALLAVNLKSLQOLD 451
Db 403 SRSIVQSTGNSQVAGAISSLFILLITVLEVCMKLPQSKVGVVTAAGVAVLVVYKL 462
Qy 452 PYLMRKSKLDCCITWVVSFLSFLSLPYGAVGVAVSVLVVVFOTQFRNGTALAQVMDT 511
Db 463 MRSIMKANRADLLIMLVFTATILNLDGLVAVIFSLLVVVTQMPHYSVLGQVPT 522
Qy 512 DIYVNPYTNRAODIOGKITTYCSPLYFANSEIF-----ROKVIK 553
Db 523 DIYDVAVETSAKEVKGKVRSSATYVFANAEPYSDALKRCGVADVDFLSQKKKLK 582
Qy 554 --TVSLOELOD---FENAPPTDP-----NNNQTPANGTSVSYTFSPDS 593
Db 583 QEOLTKLOLQKEEKLKROAASPKGASVINVTSLSDMRSNVVEDCKMKQVS---SGDK 638
Qy 594 SSPAOSPEPAEAEPGESDMLASVPPVTFHTLILDMGVSFVDLMGIRKALAKLSIYTG 653
Db 639 MEDTANGQEDSKAP-DGSTLKALGLPPDPDFHSLILDGALSFVDYVCLSKLNFHDFR 697
Qy 654 KIGKVFVFNINIAQVYNDISHGVFEDGSLCKHVPFSDIHDAVYFAQANADV 706
Db 698 EIEVEVYMAACHSPVVSQLEAGHFF-DASITKKHLFASVHDAVTFALQHPRPV 749

```

```

RESULT 3
Q9EPH0 PRELIMINARY; PRT; 744 AA.
ID Q9EPH0;
AC Q9EPH0;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PRESTIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Ludwig J., Oliver D., Frank G., Kloecker N., Gunmer A.W., Fakler B.;
RA "The reciprocal electromechanical properties of rat prestin: the motor
RT molecule of rat outer hair cells."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ303372; CAC21555.1;
DR InterPro: IPR002645; STAS.
DR Pfam: PR01740; STAS.1;
DR Pfam: PF00916; Sulfate_transp.1;
DR SEQUENCE 744 AA; 81278 MW; E49E842CF7A3CD58 CRC64;

```

Query Match 32.5%; Score 1255.5; DB 11; Length 744;
 Best Local Similarity 35.4%; Pred. No. 1.e-70;
 Matches 265; Conservative 163; Mismatches 267; Indels 53; Gaps 11;

QY	7	RYVVDRAAYSLTLEFDEFEKKDR-TYVVGKRLNAFSCSAKIKAVFGLPLVLSMPLPKY	65
Db	15	KYLVERIPFSHPVQOEKRLHKDKATBDSIGKLRQAFCFTCKKRNITIMPLPTTKMLPAY	74
QY	66	KIKDYIIPDLLGLSGSSIQVPGQMAEALLANLPAVNGLYSSEFPPLTYFFLGGVHOMVP	1255
QY	126	GTFVAFSILVGNICLOLAPES-KQVYNNATNEBYVDTAAMEARLVSATLACRATIQ	184
Db	135	GPEFVAFISLMIGVAVNRVLPVDIVIPGVNATN-----GTEARDALIRKVAASVTLSSGIIQ	190
QY	185	MGLFMOFGFAIYLSSESTFIRGFTAGLQILLSVLKTYIGLTFPSYTGSGSYFFPDI	244
Db	191	FCLGVCNRFGEPAIYLTPELVGRFTTAANAHVFTSMKLYLEGVATKRSGLFSFVYSYAV	250
QY	245	CKNLPHNTNIASL-----IFALISGAPLVLYKELNARYMHKIRREPIPTMIYVVAATAISG	299
Db	251	LQNKNNINVCSSIGLWVFGLLGG-----KEFEREKELPAIPYLPFEFAVWNGTISA	305
QY	300	GCKMPKRYHQIYVEIQGFPTPSPVYSQMKDIMGIAFSLATYSYINLAKRTLANKH	359
Db	306	GFNLHESYSDVVNVTPLPGLGLPPRANPPTSLFHLVYDAIAIAIVGFSVTTISMAKTLANKH	3655
QY	360	GVDVDSNQEMIALCNSNFGESEFKIHYICALSVTLAVDAGGKSQVASLCVSLVWITM	419
Db	366	GLOYDGNDELALGICNSIGSLFOTFSTISGSLNSLYOBEGTGKTTQAGCLASIMILVI	4255
QY	420	LVLIQIYLYLPKRSYVAGLAIANLKNLSKOLTPDPYLYMRKSKLDDCIIWVSFLSSFSLP	479
Db	426	LATFLEFESTLPQAVLSAIVINLKGMEFQSDLPFFWRTSKIELTILWTLTFVSSLEFLGD	4855
QY	480	YGVAVGAFAVLVVPDQTFNGFALAOVMDTIDIVMPKRYNNRQDIOGKIITYCSPLX	539
Db	486	YGLTVAIYIALIYIRYQSPSYTLVQLPQPTDVIYIDIDAEYKEKLPGRKIQINAPY	5435
QY	540	FANSEIFROKVIANT-----VSILOEQDEPNAPPTDPNNQTPANCTSVYI	587
Db	546	YANDLYSSALKRKTGVNPALIMGARRKAMKRYAKEVNA-----NINATVVKVDAE	598
QY	588	TFSPDSSSPAOSE-----PPASAEAPRGPSSMLASVPEPYTFHLLIDMSGSFVDMGI	642
Db	599	VDGENARKPEEDEDVKKFPPIVITTT-FPELEORFLPOGENIHRVILDFQVNMDSVGV	657
QY	643	KALKMLSTYKIGVKEFLVNIHQVYNDISHGCVFEDGSLCKCHVPEPSIHDAVLFQAN	702
Db	658	KTLAGIYKEIYDVGIIYTLACSGAQQVNDLTSNRFENPALK-ELLRHSIHDAVL-----	7111
QY	703	ARDYTPGHNFQAGFADAEALSYLSEEDI	730
Db	712	-----GSQVREMAAEDQETTVLLPQEDM	733
RESULT	4		
Q9JKO2			
AC	09JKO2	PRELIMINARY;	PRT: 744 AA.
DT	01-OCT-2000 (Tremblrel_15, Created)		
DT	01-OCT-2000 (Tremblrel_15, Last sequence update)		
DT	01-JUN-2001 (Tremblrel_17, Last annotation update)		
DE	PRESTIN.		
GN	PRESTIN.		
OS	Meriones unguiculatus (Mongolian jird).		
OC	Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Gerbillinae;		
OC	Meriones.		
OX	NCBI_TaxID=10047;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20279283; PubMed=10821263;		
RT	Zheng J., Shen W., He D.Z.Z., Long K.B., Madison L.D., Dallos P.,		
	"Prestin is the motor protein of cochlear outer hair cells."		

RL Nature 405:149-155(2000).
 DR EMBL: AF230376; AAF1715.1; -
 DR InterPro: IPR002645; STAS.
 DR InterPro: IPR001902; Sulfate_transp.
 DR Pfam: PF01740; STAS. 1
 SO Pfam: PF00916; Sulfate_transp. 1.
 SO SEQUENCE 744 AA: 8148 MW: 1CDP6589DADACBCC CRC64

Query Match	32.1%;	Score 1242.5;	DB 11;	Length 744;
Best Local Similarity	35.4%;	Pred. No. 9.2e-70;		
Matches 263;	Conservative 159;	Mismatches 278;	Indels 43;	Gaps 10

QY	7	RYVVDRAAYSLTLEDDDEEKKDR--TYPYGEKJLNAFCSCKAKIYAVFGLLPVYLSMPLPK	65
Db	15	KYHERPILFHSHPVQERLHMYDKXSESIGDKLAKAFCTPKKINILYMLPIITKMLPAY	74
QY	66	KIKDYIIPDLGLSGSGSIQVPGCMARFALLANILPANGVLYSFFPLLTYFFLPGVHOMVP	125
Db	75	KKEFEYVIEDLVSGISGTGVLQIPGLIAFMAAPVPEGLVSSEFPVLMYCGFFGSRHAI	134
QY	126	GTFPAVSLVYVNICLOLAPES--KQOVFNMMANESYVDFPAMEARLHVSAFTACTPFIQ	184
Db	135	GPFAVISMIGVAVRLVPPDDIYIPGVYNATN---GTEADALRVVAAHVSLLSGLTIQ	190
QY	185	MLGLFMGCGFAIYLSSESFIQFMTAAGLQILISLVKLYIGLTIIPSYGSGSIVFTFIDI	244
Db	191	FCLGVCRCRGFAIYLTBEVLVNGFTTAAVAHFTSMKLYRGVKRRKRSGLIFSIVYSHAV	250
QY	245	CKNLPHNINIASL----IPALISAPFLVLYKELNARYMHKIRPIPIEMIVAVVA7AISC	299
Db	251	LQNVKNLWVCSLGVLMWFGLLGSG----KEPNERKEKLPAPILPELFVAVWGCTISA	309
QY	300	GCKMKPKYHMDIYVEIQNGFPFPSPVYVSOCKMDIGAFSAIYSYINLMMGRTLANKH	359
Db	306	GFNLHESYSVQVVGTLPLGLLPPAPDTSLEHLVYVDAIAIAIGSVYTTSMATLANKH	366
QY	360	GVDVDSNOEMIALGCSNFFGSPFKIHWICALSVTLAVDAGKRSQVYASLCVSLVNAITM	419
Db	366	GQYVDNGQELIALGICNSIGSLFQTFSSICLSLSRSLVQEGTGKTLQAGCLASIMILLVI	425
QY	420	LYVLGIYLPLEKSVYLGALIANVNLKSLKQLDPIYLYMRKSLDDCIIWVSFSLFSLP	479
Db	426	LATGFLFEESLQOAVLSAIVYINLGMENQFSDLPFEWTSKIELTILMLTTFVSSLFGLD	485
QY	480	YGVAAVGVAFSVLVYVFOQGFQNFNGALAOVMDTJIVNPKTYNRADODLOGIKIITYCSPLY	539
Db	486	YGLTAAVIAIALLIYTYRKQSPSYKVLGDLPTDVIYIDDAIEVKELPGLKIIPIINAPIY	545
QY	540	FANSEIFROKVIYAT-----VSILOQODEENAPPTDPNNQTPPANGTSVSYI	587
Db	546	YANSDLYSNALRKRTGVPNALIMGARRKAMKRYAKEVGN--N1ANAAVYVKGDEVGDN	603
QY	588	TFSPDSSSPADSEPPASAEABGEPSDMLASPPRYTFTLTLIDMSGSFYVDLMKIRLAK	647
Db	604	ATKPEEEDDEVKYPPIYIKTT--FPPELOREFMPQRENVHTIILDPYQVFNDSVGVKTLAV	667
QY	648	LSSTYGRKGAVVFLVNIHAQVYVNIOSHGVGEEDSLCKEHVFSIHAVALFPAQANARDVT	707
Db	663	MKKEKGVGVGIYVYIYLAGCSPQVYVNDLTRRRFFENFALK-ELLFHSIHDAVL-----	711
QY	708	PGHNFQAPGADAEISLYDSEEDI 730	
Db	712	GSYHREAMAEADSEASAPPOQDM 733	
RESULT	5		
Q099NH7			
AC	Q099NH7	PRELIMINARY;	PRT; 744 AA.
DT	01-JUN-2001	(Tremblrel. 17, Created)	
DT	01-JUN-2001	(Tremblrel. 17, last sequence update)	
DT	01-JUN-2001	(Tremblrel. 17, last annotation update)	

DE PRESTIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=ORGAN OF CORTI;
RA Adler H.J., Meritt R.C. Jr., Belyantseva I.A., Kachar B.;
RT "Mus musculus organ of Corti mRNA for prestin."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY024359; AAG5999.1; -
SQ SEQUENCE 744 AA; 81380 MM; 56E842748288E96F CRC64;

Query Match 32.1%; Score 1242; DB 11; Length 744;
Best Local Similarity 36.5%; Pred. No. 9.9e-70;
Matches 262; Conservative 154; Mismatches 260; Indels 42; Gaps 10;

QY 7 RYVVDRAVYSLTLPDEFEKDR-TYPYGEKLNAPFCSAKIKAVFGLPLVLSMLPKY 65
DB 15 RYVVERPFSHPVLQERLHVMDKVTESIGDKQAFTYTPKKIRNIIIMFLPTKMLPAY 74
QY 66 KIKDYIIPDLGLSGSGSIQVPGMAFALLANPAVNGLYSFFPLTYFFLGVHOMP 125
DB 75 KFEKVLQDLVSGISTGVLOLPQGLAFAMLAAPVFGLYSFFVIMVCFPGTSRHISI 134
QY 126 GTFAYISLVGNICLOLAPES-KFQVFNATNESYVDTAAMEARLHSATLACTLTATIQ 184
DB 135 GPFAYISLMIGVAVRLVPDDIVIRGVNATN---GTARDALNKKAMSTLLSGITQ 190
QY 185 MGLGFMQGFVAIYLSSEFIRGEFTAGLIQILISLVKYLFGITLTPSYGPGSIVETFDI 244
DB 191 FCLGRCRGFAIYLTPELVRGFTTAAAVHFTSMKLYLFGVKTARGSGIFSIVSTAV 250
QY 245 CKNLPHNTIASL-----IFALISGAFVLVKELNARMKIRFPTEMYVVAFTAISG 299
DB 251 LQNVKNLVNCSLGVGLMFGLLGG---KEFNERFEKEKPARIPLEFFAVMGTGISA 305
QY 300 GCKMPKRRHMOIVGEIQRGFPPTPVSPVVSQMKDMGTAFSLAIYVYINLAMPRTLANKH 359
DB 306 GPNLHESIVDVVGTLPGLPRLPANPDTSLFHLVYVDALAIAYGFSVTISMAKTLANKH 365
QY 360 GYVDVDSNQMIALGCSNFGSFFKIHVICALSVTLAVDAGAGKSOVASLCSLVVMTM 419
DB 366 GYVDVDSNQMIALGCSNFGSFFKIHVICALSVTLAVDAGAGKSOVASLCSLVVMTM 419
QY 420 LVLGILYLPKRSVGLAIANLKNLSKQLTDPYILMRKSKLDCCIWVVSFLSFLSLP 479
DB 426 LATGFLFESLQAVLSAIVNLKGMFQSDLPFEMRTSKIEPTIMLTTEVSSFLGLD 485
QY 480 YGVANAVFSLVAVVFOFRNGVALAQMVDIYVNPRTNRRADIOGIRITTCSPLY 539
DB 486 YGLTAVTIALTYIKTQSPSYKVLGQLPDVTIDDAIEVEVLEIGIKIFQINAPLY 545
QY 540 FANSEIFROKVIATY-----VSLOEQDFENAPPTDPNNQTPANGTSVSYI 587
DB 546 YANSOLYSALAKRKTGVNAPALIMGARAKAMKRYAKEGNA-----NVANATVVKVAAE 598
QY 588 TFSPOSSSPAQSE-----PPASAEAPGEPDMLASVPPVTFPHLLIDMSGVSEVDLMTI 642
DB 599 VDGEMATPEEEDDEVKPPPIVIKTT-PPETLORELPOGEVNHVILDFTOVNEFDSGV 657
QY 643 KALAKLSLTYGKIGKVVFLVNIHMOVYNDISHGVFEEDSGLECKHVPFSIDAVLFAQ 700
DB 658 KTLGAIYKEYDVGIVYLAGCSPPQVNDLTRNNEFENPALK-ELLFHSIDAVLGSQ 714

RESULT 6
Q9RI55 PRELIMINARY; PRT: 780 AA.
AC Q9RI55;
DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE PENDRIN.
CN SLC26A4 OR PDS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BA1B/C;
RX MEDLINE=99380587; PubMed=10449762;
RA Everett L.A., Morsli H., Wu D.K., Green E.D.;
RT "Expression pattern of the mouse Ortholog of the Pendrin Syndrome Gene
(Pds) Suggests a Key Role for Pendrin in the Inner Ear."
RL Proc. Natl. Acad. Sci. U.S.A. 96:9727-9732(1999).
DR EMBL: AF167411; AAD51617.1; -
DR MGD: MGI:1346029; SLC26A4.
DR InterPro: IPR002645; STAS.
DR InterPro: IPR001902; Sulfate_transp.
DR Pfam: PF001740; STAS.1.
DR Pfam: PF00916; Sulfate_transp.1.
SQ SEQUENCE 780 AA; 85686 MM; 4988B48F057BF38 CRC64;

Query Match 30.6%; Score 1185; DB 11; Length 780;
Best Local Similarity 35.1%; Pred. No. 3.8e-66;
Matches 272; Conservative 147; Mismatches 297; Indels 60; Gaps 13;

QY 8 YVVDRAVYSLTLPDEFEKDRTPYGEKLNAPF-----RCSAKIKAVFGLPLVLSMLP 63
DB 20 YTVSPVYSELAFQ---QQERRRLPERRTLRSLARSCSKRAFGVAKTLPLDLMP 76
QY 64 KYKIDYIIPDLGLSGSGSIQVPGMAFALLANPAVNGLYSFFPLTYFFLGVHOMP 123
DB 77 KRYAKEMLLSDIISGVSTGLVGTLOGMAVALLAAPPVFGLSAFPLTYTFVGTSMHI 136
QY 124 VPGFAYISLVGNICLOLAPESKQV---FNATNESYVDTAAMEARLHSATLACT 180
DB 137 SVGPFPVYSLWVGSVLSMAPDDHFLVPSNGSALNSTLDTGTDARVILASLTLLV 196
QY 181 AIQMGCLFMQGFVAIYLSSEFIRGEFTAGLIQILISLVKYLFGITLTPSYGPGSIVFT 240
DB 197 GIQVLFGLQIGFVRLADPLVGGFTTAAAFQVLSQILVNVSTKNNGILSIYIT 256
QY 241 FIDCKNLPHNTIASLIFALISGAFVLVKELNARMKIRFPTEMYVVAFTAISG 300
DB 257 LIETQNTGDNIMADFIAGLLITIVCMKAVKELNDFKRRIPVPIPIEIVYITTAISYG 316
QY 301 CCKMPKRRHMOIVGEIQRGFPPTPVSPVVSQMKDMGTAFSLAIYVYINLAMPRTLANKH 360
DB 317 ANLEKNYAGIYKSLPSGFLPPLVPSVGLFSDMLAASISIAVNAIYVSGKYATFKHD 376
QY 361 YVDVDSNQMIALGCSNFGSFFKIHVICALSVTLAVDAGAGKSOVASLCSLVVMTM 420
DB 377 YVIDNQEPIAFGISNVFSGFCFVATTALSRVQESTGKIQVAGLISIAVIVMAIV 436
QY 421 VLGITLYLPKRSVGLAIANLKNLSKQLTDPYILMRKSKLDCCIWVVSFLSFLSLP 480
DB 437 ALGRLEPLOSYSVLAAYIANLKGFMKQVCPRLMKONKTDATVATWETCMTSITLGLD 496
QY 481 GVAAGVAFSLVAVVFOFRNGVALAQMVDIYVNPRTNRRADIOGIRITTCSPLY 540
DB 497 GLLAGLLEFALLTVLRVQFPFMNGLGSPSPDIXKSTIHYKNLEPBEVKILRRSSPTFY 556
QY 541 ANSEIFROKVIATY-----VSLOEQDFENAPPTDPNNQTPANGTSVSYI 587
DB 557 GNVDGFK-KCINSTYGFDAIRYNNRKLARRIOKLKGLRATKNGIISDSSNN-- 613
QY 588 TFSPOSSSPAQSEPPASAEAGCE-----PSDMLASVP-----PFTPHLLIDMSGVSEVD 638
DB 614 AFEDE-----DVEPEELNPTKIEIQVDWNSLEPVKVVNPKVISHVLVDCGAVSFLD 669

OY	639	LMGKALATLSTSYTKGVVFLVFNJHAOYNDHSHGEDEGSJECKHVPSIHDAVLF	698
Db	670	VVGFRSLMAYKEQRIDVAVNYFALLQDDYLEMECCGFEDD-NRKDRPFLTYADALIH	728
OY	699	AQANARDVTGHNPGAPGADELSTLYDSEEDIRSYWD---LEOFNFGSMFAETLL	750
Db	729	LQNQVKS-----REGQDSLTLETARIRDKCDPLDMLEAEK-----NAEEL	768
RESULT	7		
ID	O9R154	PRELIMINARY;	PRT; 780 AA.
AC	O9R154;		
DT	01-MAY-2000 (Tremblrel. 13, Created)		
DT	01-MAY-2000 (Tremblrel. 13, Last sequence update)		
DT	01-JUN-2001 (Tremblrel. 17, last annotation update)		
DE	PENDRIIN.		
GN	pds.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.		
OX	NCBI_Taxid:10116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=SPRAGUE-DAWLEY;		
RX	MEDLINE=99380587; PubMed=10449762;		
RA	Everett L.A., Morsli H., Wu D.K., Green E.D.;		
RT	"Expression pattern of the Mouse Ortholog of the Pendred Syndrome Gene		
RT	(pds) Suggests a Key Role for Pendrin in the Inner Ear.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 96:9727-9732(1999).		
DR	EMBL: AF167412; AAD51618.1; -		
DR	InterPro: IPR001902; Sulfate_transp.		
DR	InterPro: IPR002645; STAS.		
DR	Pfam: PF00916; Sulfate_transp; 1.		
DR	Pfam: PFO1740; STAS; 1.		
DR	PROSITE: PS01130; SULFATE_TRANSF; UNKNOWN_1.		
SQ	SEQUENCE 780 AA; 85714 MW; DAOCDB7496BD0535 CRC64;		
Query Match	30.6%; Score 1184; DB 11; Length 780;		
Best Local Similarity	36.3%; Pred. No. 4,4e-66;		
Matches 264; Conservative 143; Mismatches 277; Indels 44; Gaps			
OY	8	YVDDRAAIVSLTFDEDEKKDRTPYVEGIKRN--AFRCSSAKIKAVFG----LIIPLVIS	61
Db	20	YAVERPVPYSIELAFQ---QORERRLPERRTRLRDLSIASCSCSRKA--FGALKALPIIDW	74
OY	62	LPKKIKDIYIPDLGLSGSIOVPOGMAFALLAMPVANGISSEFPILTYFFLGCVH	121
Db	75	LPKRIVEWLSDIIISGVSTGLVGTGGMAVALAAVPVOYGISAFFPILTYFEVGTSR	134
OY	122	QMVGFTFAVISIILVNGICLOLAPESKPV--FNNAT-NESYVDTAAMEERLVSATLAC	178
Db	135	HISVGPPPVYSIAMGSVVLSMAPRDDHFVYPGSGSLINTTTLDTRGRDARARLLASTLTL	194
OY	179	LTAIIOMGLGFMORGFVAIYLSESFINGFWTAAGILOILSVLKTIYFGLTIPSPTVGPSIV	238
Db	195	LVGIITOLVFGLQIGFIYRYLRADPLVCGFTTAAPQVLSQIKIYLVNSTKNVNGVLSII	254
OY	239	FTFIDICKNLPHNTIASLIFALISGAFLVYKELNARKMKIRFPLPTEMIYVVAVTAIS	298
Db	255	YTLIEIFONIGNDNPIADFINGLLTLIIYCAVAKELNRDFHKHIPVPIPIEVITIATAIS	314
OY	299	GGCAMPKYHHOAIYGEIORGFPPVPSPVVSOMKDMLGTAFSLAIYSVINLAMGRTLANK	358
Db	315	YGANLEANTNAGVIKTSIPSGFLRPVLPVSGLFSDMLAASSTIAVVAIALVAVSGVAVTK	374
OY	359	HGYDVDSNOEMIALGCSNFSGSEFKIHVICALSVTTLAVDAGAGKSOVASLCSLVMIT	418
Db	375	HDYIIIDGNQEPIARGISNVFSGPFSCVATLSTRVAGDESIGKQVAGLLSATIVWA	434
OY	419	MLVIGCIYLYPLPKSVLDALIAVNKLNSKOLDPYTLVMRSKIDCCIWVVSFLSSFETSL	478

Db	435	IVALGKLEPELOKSVLAAYVIANLKGMFQVOCVPRLMKQNKTDVAIWMFTCLMSITLGL	494
Qy	479	PYCAVAVAFSVLVVVFOTQFRNCGYALAOYMDTDIYNPKTYNRAQDIOGKITTCYSL	538
Db	495	DLGIAGLALFGLLTVLRVYVFPSPNMGSGVSPSTDIVYSIKTHYKMLEBEQVKKILRESSPI	554
Qy	539	YFAASELFPKQVIAKVY-----SLOELODFENAPPTDNNNQTFANGTSYS	585
Db	555	FYGVNDGPK-KCAVSTVGFDALIRYNNKRLKALRIQLOLIRKQGLRATKNGIISDVGSNN	613
Qy	586	YTFSPSSSPAQSEPPASAEPGE-----PSDMIASVP-----PFVYFHTLLIDMSGVSF	636
Db	614	--AEPE-----DVEPEBELDIPKKEILEIQYDWMSELPRVKNPKYHSLVLDGCAVSF	667
Qy	637	VDLMGIALAKLSTSYKIGVYFLVNIHAQVYNDISHGVFEDSGLECKHVPSPSHDAV	656
Db	668	LDVAVGSLRMYKEFORIDVNYFYALLDDVLEKMEGCGFFDD-NIRKDRFELYVDAL	726
Qy	697	LFAQANR 704	
Db	727	LYLQNOAK 734	
RESULT 8			
Q9WVC8			
AC	Q9WVC8	PRELIMINARY:	PRT: 757 AA.
DT	01-NOV-1999 (TREMBLrel. 12, Created)		
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE	DOWN-REGULATED IN ADENOMA PROTEIN.		
CN	SIC26A3 OR DBA.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.		
OX	NCBI_TaxId=10099;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=EV8/N; TISSUE=COLON;		
RA	Meivin J.E., Park K., Richardson L.A., Schulteis P.J., Shull G.E.;		
RT	Mouse Down-Regulated in Adenoma (DBA) is an intestinal Cl-/HCO3-		
RT	Exchanger and is upregulated in Colon of Mice Lacking the NHE3 Na+/H+		
RT	Exchanger.";		
RL	J. Biol. Chem. 0:0-0(1999).		
DR	EMBL: AF136751; AAD2784.1; -		
DR	MGI: 107181; SIC26a3.		
DR	Interpro: IPR002645; STAS.		
DR	Interpro: IPR001902; Sulfate.transp.		
DR	Pfam: PF01740; STAS: 1.		
DR	Pfam: PF00916; Sulfate.transp: 1.		
DR	PROSITE: PS01130; SULFATE_TRANSFP; UNKNOWN_1.		
SO	SEQUENCE 757 AA; 83589 MW; 32BJAC648BE7A07 CRC64;		
Query Match 29.7%; Score 1147.5; DB 11: Length 757;			
Best Local Similarity 34.7%; Pred. No. 8.2e-64;			
Matches 254; Conservative 149; Mismatches 261; Indels 67; Gaps			
Qy	7	RYVVDRAVYSLTLDDEFEKKDRTPYG-EKLRNAPFCSSAKITAAVFGLLPYLSMLPKY	65
Db	8	QYVAVRPVYSTRKTGEFFKTHRHKKFLDLKCCSSCKKAKIALSLSPRIASMLPAY	67
Qy	66	KIKYIITPDLLGSGSGSIOVPGCMAPALLANLPVANGVSSFPPLITLYFFLGVHOMVP	125
Db	68	KIKEMLLSDIYSGISGLGAVAVLQGLAFRLALVNIIPAVGLIAAFPVITTYFFLGSRHSIV	127
Qy	126	GTFVAVISLVGNICLOLA--PESKFOYFNNAF--NESYVDTAAEAEERLHVASATLACTJAI	182
Db	128	GPPEVLMSMWGVVTVRVSDPNASSELSSSTENDSFIE-----EKVMAASVTVLSGI	181
Qy	183	IQMLGFMQGFVAIYLSESTIRPFMTAAQGLIILSVLKITFGLTITSYGPGSIVTTFI	242
Db	182	IQMLGFMQGFVAIYLSESTIRPFMTAAQGLIILSVLKITFGLTITSYGPGSIVTTFI	241

OY	243	DICKLPTNTNIAFLIASAFILVLEKLNARMHTRIPPEMIVVVAVATASGGC	302
OY	243	: : : : : : : : :	
Db	242	SVFIOIKNTIADLVTSIIILVVFVKREINQRKSLLPPPIPELIIMTYIANGVSIGN	301
OY	303	MKKYHMOIVGEIORGFPTPVSPVSGWKMDITGFSLAISVYINLAMRTLANRHYD	362
Db	302	FEDRGVAVVNMSIGFOPPITPSVEVFODTIDGSFGIAIVGVAFSVASVSLKYDP	361
OY	363	VDSNEMALCALCSNPFGSEFFKHVICALLSTVLAVDAGGKSGOVASCISLYVMITMLYL	422
Db	362	IDGNELIALGLVSNIFTCFAFGFAGSTRLSRSGVESTGKTQVAGLSAVILIYVAI	421
OY	423	GITYLPLEKSYVALGAIIVNLKNSLKOLDTPPYLMRKSKLDCCICWVSFSLSLPYGV	482
Db	422	GFLLOPPLORSVALAIALNLKGMLMOEAEIRLTKMKKCYCLIMTFEFAIVLGGLGL	481
OY	483	AVGVAFSVLVVPOTQPFNGVALAQWDDTDIYNPKNYRNAODLOGIKITTYSPLYEAN	542
Db	482	AASVAFOLLITTVFRTOFPKSCSTLANVGRSNYYKKKNNAEYVEEGVKIRCPSPIFYAN	541
OY	543	SEIFPROXY-----AKTVSILOELOQ-----DEFENAPPTD	571
Db	542	IGFEFKOLIDAVNGFSPULRLKRNNKALKIKRIQLRGILQMPKGICTSDGFKDDEL	601
OY	572	PNNN----QTANSTSVSY-ITFSPDSSSPAOSPPASAERGPSPMLASVPEFYFHT	626
Db	602	DNQIIEILDOPINTTDLDFPIDMWGDL-----LWITTPKLSHS	641
OY	627	LILDMGSAFPDNLGAIKANLASSTYGKIGVKEVLVINHAOVYNDISHGCVFEDGSLECK	686
Db	642	LILDFSAAVSFDVSSMRGLRTIIGEFIRIKVDYIVGTDDDFDKLARCFEFD-EVTDS	700
OY	687	HVFPSHDVYL 697	
Db	701	IPELLIHDAVL 711	
<hr/>			
RESULT	9		
O9BEG8			
ID	O9BEG8	PRELIMINARY;	PRT; 734 AA.
AC	O9BEG8:		
DT	01-JUN-2001 (TREMBLrel_17, Created)		
DT	01-JUN-2001 (TREMBLrel_17, Last sequence update)		
DT	01-JUN-2001 (TREMBLrel_17, Last annotation update)		
DE	DTD SULFATE TRANSPORTER.		
OS	Bos taurus (Bovine).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;		
OC	Bovidae; Bovinae; Bos.		
OX	NCBL_TaxID=9913;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Baumgartner B.G., Kriegesmann B., Brenig B.;		
RT	"Cloning and characterization of the bovine DTD sulfate transporter		
RT	(dbtbsp) gene."		
RL	Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AJ223615; CAB69640.1;		
SO	SEQUENCE /34 AA; 81540 MW; D5224A27EA5D691E CRC64;		

Query Match	26.3%	Score 1016.5	DB 6	Length 734
Best Local Similarity	32.8%	Pred. No. 1.2e-55		
Matches 231	Conservative 137	Mismatches 286	Indels 51	Gaps 9

57
Db 61 EKSDNNEKFKIKLEKSCQSSFRKANTJFGLFVLQMLPKYDKLNNILGDWMSGLIYG 1200
83 SIOYDQGAFFALLNLPVANGVLSFSPFLLTYFELGWHQWVGTFAVISILVGNICLOL 142
Db 121 ILVPOGSLAYLLAGQEPETIGLYLVSFPASLISYFLLGTSRHI SWGIFICLDMGEVADRE 1800

0Y	143	APESFQVFNNAATNSY-----DTAMEERLHVSAITLCLAIIONMGJFQFQFEVA	196
Db	181	LYIAGTDTVHAASNESSLVNODISOKTCDRCSCIYIIIVGSTYFVAGVQVAMGFFQVGSV	240
0Y	197	ILYSESPFRGFMTAAGLOILISLVKTYIFGLTIPSTYGPGSIVTFEIDCNLPHTIASL	256
Db	241	VLSDALLGVEGTASFTILTSQYKLLGLISLPSRAGVGSILTTMLHVFNRIRKTNICDL	300
0Y	257	IFALISGAFVLVLVELNARYMHKIRPPTPTMIYVVAVATISGCCMKPKRTYHMOIYEIQ	316
Db	301	ITSLCLLTVLLPTKELNERFKSKLAPIPVELFVIVATLASHGKLNKNGTSTIAGHIP	360
0Y	317	RGEFPPVPSVYOSMKDMTGTFASLAIYSVYINLMAGRTIANKHGYDVSNOEMIALGCSN	376
Db	361	TGFMPKAPDMNLIPRAVADIAITAIIGFALTYSLEMFPAKHGTYTAKADEMATISFCN	420
0Y	377	FFGSEFFKIHVICCALSVLADVAGGAGSKQVASLCSVSLVMVTIMVLGILYLLPPLKSVLGA	436
Db	421	IIPSEFHCFITTSALAKTLVKESTGCGQVGVMTALLVLLVIAPIPLFSLOKSVLGV	480
0Y	437	LIAVNLKNSLKQLTDPYLYLAKRSKLDCCIWVVSFLSEFISLPYGAVGAFSVLYVVFQ	496
Db	481	ITIVNLGALCKRFDLPOMMRISHMDDYIMFVTMLSSALSTEIGLLTGCFCSMFCYLIR	540
0Y	497	TOFRNGYALAAVMDIYVNRKPTNRNODLOGICIIITCYSPLYANSEIROKYATIAS	556
Db	541	TOKPRASLLGLVEDESEVESASAKNLQAKGIIIFRVAFLYVNVNEYKSVLYKTTLN	600
0Y	557	LOELOODEENAPPDNNNNQTPANGTSVSYITFSPDSSSPASOSEPPASAEAPCEPSDMA	616
Db	601	-----PVLKRAAQKAKKAKRKIKRETYIP-----SGIQDE--VSYQLSHDP-----	638
0Y	617	SVPPETVTEHTLLIDSGVSEVDLKGICALAKSLSTYGGIKGVKFLVNIHAQVINDISHGG	676
Db	639	-----LEEHITVIVDSCAIOFEDTGAIHITLKEVRDYDEAVGIVOLYLAOCNPSVRDSLARGE	693
0Y	677	VFEDOSLECKH-----VFPSITHDAVLPAQ-----ANARDYTPGHNF	712
Db	694	Y-----CKKDEENLEFYSITYEMTAEPRASQONOKERHINPQPNF	731

RESULT	10			
Q9GJY3				
ID	Q9GJY3	PRELIMINARY;	PRT;	734 AA.
AC	Q9GJY3;			
DT	01-MAR-2001 (TRENBLREL, 16, Created)			
DT	01-MAR-2001 (TRENBLREL, 16, Last sequence update)			
DT	01-JUN-2001 (TRENBLREL, 17, Last annotation update)			
DE	SULFATE TRANSPORTER.			
GN	ST.			
OS	Ovis aries (Sheep).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eute			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora			
OC	Bovidae; Caprinae; Ovis.			
OX	NCBI_TaxID=9940;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Kriegesmann B., Baumgartner B.G., Deppe A., Brenig B.,			
RT	"Sequence of the ovine sulfate transporter gene.";			
RL	Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases			
DR	EMBL, Y18558; CAC20729.1; -			
DR	InterPro: IPR002645; STAS.			
DR	InterPro: IPR001902; Sulfate_transp.			
DR	Pfam: PF01740; STAS: 1.			
DR	Pfam: PF00916; Sulfate_transp: 1.			
DR	PROSITE: PS01130; SULFATE_TRANSF: 1.			
SO	SEQUENCE 734 AA; 81439 MW; 35EABDFB4F2F3B36 CRC64;			

Query Match	25.9%;	Score 1001.5;	DB 6;	Length 734;
Best Local Similarity	33.2%;	Pred. No. 1.1e-54;		
Matches 229;	Conservative 132;	Mismatches 281;	Indels 47;	Gaps 8;

QY	25	EKKRTPP---	VEGELTRAPFSSCKIKAVVEGGLLPVLSMPKXIKOYII	PDLLGGISG	82
Db	61	EKSDNNLKRYIKLEKSCCOSTKAKNTT	FGFLPVLOMLPKTDLKNILGDMMSGI	IVG	120
QY	83	STOYPOGMAEALLANLPAVNGLSYSSFP	LLTYFFELGSHOMVGTEFAVISILVGNIC	LIOL	142
Db	121	ILVPOGIAVSLLAGOEPRTGYLTSPFAS	ILITFLGSRHISVIGRISLCLMIGEVDRE		180
QY	143	APESKFOVFNNAINESVY-----	DTAMEAERLHNSATLACTLAIIO	MGFGMEFVA	196
Db	181	LYINGDTVAHAASNESSLVNOMSNOT	DRCSCAIIVGSIYTFPAVAGYQVAMGE	FFGYGFW	240
QY	197	ITYSESRIGRMPTAGQOILISVLYKIF	GLTIPSYNGPSIVTFPFIDICNLRHTNTASL		256
Db	241	YISDALLGEGVTGASTTILTSQVKYLL	GLISLPRSGVGSLITTWIIIFNNIKHTN	ICDL	300
QY	257	IFALISGAFVLVVELNARVYMKIRP	IPLEMIVVAVATISGCGKPKRYHMOI	VEIG	316
Db	301	ITSILCLLVLLPTELNERKSKAKAIP	PELPPVVAATLASHHGKISEKTSIAGHIP		360
QY	317	RGFPYSPVVSQMKDMIGAFSLAIVY	INLAMEGTTLANKHGVDYSNOEMIALG	CSN	376
Db	361	TGFMPKAPDMNLIPIVAVDATAI	IGFAITVSLSEMPAKKHGYTVKANQEM	VAIGFCN	420
QY	377	FFGSEFKIHYICALSTLAVDAGCGS	QVASLCVSLVNMITMLVGLIYIPLKSYLGA		436
Db	421	IIIPFHSFTTSAALATLVKESGCG	TCOTQSGWNTALVLLVLLIAPLFFSL	OKSVLGV	480
QY	437	LIAVNLKNSIKOLDTPPYLYMKRSK	LDCCIVWSFSLSEFSLDPLGYAVNGAF	VLVVPQ	496
Db	481	ITYNLGALCKFKDLPOMKMSIMD	IVIMFTMLSSALISTEIGLITGVCFS	MFVILR	540
QY	497	TOPFNGYALAOVMDPTIVNPKTYN	RADODIGIKIITYCSPLYFANSEIR	PROKVIATVS	556
Db	541	TQKKAISLGLVESEVFEFSASAKN	LNQASGIKIIFPVAPLVYVKNKYSV	LKXTLN	600
QY	557	LQELQDPENAPPTDPNNNOTPANG	ISVSYTSSPDSSPAQSEPPASABE	PEPSDMA	616
Db	601	-----PVLVYAAROKAKKAKKIK	RETV---TLGSIODE--VSVOLSDP	-----	638
QY	617	SVFPFVTFHTLLDMSVSPFDMG	IKALKLSTGYKIGCVKPLVNIHMOV	NDISHGG	676
Db	639	-----LEHTTIVDCSAIOFLDTAG	HTLKEVRDYPATIGIYVLLAOCN	SVHDSLRGE	693
QY	677	VFEDGSLECKH-----VFPSI	HDAVLEP	700	
Db	694	Y-----CKKDEENLIFYSE	EMTPAE	715	
RESULT	11				
Q9BGH1					
AC	09BGH1	PRELIMINARY;	PRT;	463	AA.
DT	01-JUN-2001	(TEMBLrel. 17, Created)			
DT	01-JUN-2001	(TEMBLrel. 17, Last sequence update)			
DT	01-JUN-2001	(TEMBLrel. 17, Last annotation update)			
DE	DOWN-REGULATED IN ADENOMA	DRA (FRAGMENT).			
GN	SIC2.6A3.				
OS	Oryctolagus cuniculus (Rabbit).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.				
OX	NCBI_TaxID=9986;				
RN	11				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=INTESTINAL MUCOSA;				
RA	Jacob P., Rossmann H., Lamprecht G., Kretz A., Neff C., Lin-Wu E.,				
RA	Gregor M., Seidler U.;				
RT	Characterization of rat and rabbit brush border membrane anion				
RT	exchange and DRA expression in rabbit, rat, and human duodenum.;				
RL	Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL, AF314819; AAK00897.1; -.				
FT	NON_TER	1			
FT	NON_TER	463			

[illegible]

Db 2 NIPPAVGLYAAFEFPVITYFFELGTSNRHISVGPPLVSMVGVVTVRVASGSD--TSPALS 58
 Qy 157 ESYVDTAAEAEERLHVSAITLACTAIIOMGLGFMQGFVAIYLSSEFIRGFMFAAGLOIL 216
 Db 59 SSSAENDSIEEKVAVASVTVLGSITQILLQIDIGFVITYLSLISGFTTAAAIHVL 118
 Qy 217 ISLVKIFGLTIPSTYGPESIVFTFDICKNLPHTNIASLIFALISGAFLVVKELNARY 276
 Db 119 VSQLFMQLQTFPAHSDPFISIKVLESVFSQIQKTNIDLVTSVILLVVFAVKEINQRY 178
 Qy 277 MHKIFPIPTFEMIVVVVVAIISGCKMPKKYHMOIVGELQREPFPPVSVSOMKDMCT 336
 Db 179 RSKLVEPIPTFELIMTVIATGISYGCNFEOREGVAVGNNSLGFQPTIPSVVEFDDTGD 238
 Qy 337 AFSLAIVSVIIMANGRTLANKGYDVDSNQEMIALGCSNFGSGFEKIHVICALSVTLA 396
 Db 239 CGCIAIVGVAFAVSASVSLKYHPIDGHQELIALGVSNITGAFKGNASTALSRSV 298
 Qy 397 VDGAGCKSOVASLCSVLYVMITMLVGLIYLPKPSVGCALIAVLKNSLKQLTDPYYLM 456
 Db 299 QESXGKTOVAGLISAVIYLIYVAIGFLQPLQKSVLAALAGLKGMLQFAEIGRLM 358
 Qy 457 RSKRIDCCIMVVSFLSPFSLPYGAVGVAFAVSVLVVFQRFNGYALAOVMDTIYVN 516
 Db 359 KKDQKIDCLIMVTFIATYLGIGLGLASVAFOLITVFRTOFPCOSTLANGRSNTYKN 418
 Qy 517 PKTYNRADIDQIKIITYCSPLYFANSEIFROKVI 551
 Db 419 KKNYADVEPEGVKIFRCPSPIYFANIGFEKOLI 453

RESULT 13

Qy 09H2B4 PRELIMINARY: PRT; 701 AA.
 AC 09H2B4;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE SULFATE/ANION TRANSPORTER SAT-1 PROTEIN.
 GN SLC26A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RX MEDLINE=20541715; PubMed=11087667;
 RA Lohli H., Kujala M., Kerkela E., Saarialho-kere U., Kestlia M.,
 Kere J.;
 RT "Mapping of five New Putative Anion Transporter Genes in Human and
 RT Characterization of SLC26A6, A Candidate Gene for Pancreatic Anion
 RT Exchanger.";
 RL Genomics 70:102-112(2000).
 DR EMBL: AF297659; AAC22075.1; -
 DR InterPro: IPR003880; Phosphopent_atach.
 DR InterPro: IPR002645; STAS.
 DR InterPro: IPR001902; Sulfate_transp.
 DR Pfam: PF01740; STAS; 1.
 DR Pfam: PF00916; Sulfate_transp; 1.
 DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_1.
 SO SEQUENCE 701 AA; 74991 MW; F37DC87F6A036676 CRC64;

Query Match

22.7%; Score 878; DB 4; Length 701;

Best Local Similarity 29.6%; Pred. No. 5.2e-47; Mismatches 209; Conservative 146; Mismatches 304; Indels 46; Gaps 8;

Qy 29 RTYPVGEKLRNFR-----CSSAKIKAVFGLLPVLSPPKIKIDYIIPDLGLGS 80
 Db 19 RORPAPGRGLREMLKARLWCSGCSYLCVRAVLQDILLPATRMLRGYRRPREVLAGDVMGSLV 78
 Qy 81 GGSIOVPGGMFAALLANLPVANGLYSSFPPLITTFYFLGCVHOMVPGTFAVISILVGNIC- 139

Db 79 IGIIIVPOAIAVSLLAGIOPISLYTSEFANLIYELMGSRHVSIGFISLCLAMGOVVD 138
 Qy 140 --LOLA--PESEFOYFNNAATNESYVDTAAM-----EAEHLHSAITLACTTAIOMGLGFM 190
 Db 139 RELQLAGDPDSODGLQPGANSSSTLNGSAMLDCGDCAIRATVATLTGYOLMGLV 198
 Qy 191 QGFVAIYLSSEFIRGFMFAAGLOILISVLKYIFGLTIPSTYGPESIVFTFDICKNLPH 250
 Db 199 RLGFSAVLSQPLDGFPMAGASVTILTSLKILLGVRLPRHGGPMVVLTMSSLRGAGQ 258
 Qy 251 TNIASLIFALISGAFLVVKELNARYMHKIRPITFEMIVVVVVAIISGCKMPKKYHMO 310
 Db 259 ANVCDDVSTVCLAVLAAKELSDRYRRRLVPTTELLIVAVATLVHFGQLHRRFGSS 318
 Qy 311 IYGEIORGEPFPPVSVSOMKDMIGTAFSLAVSVIIVNAMGRTLANKRGYDVDSNOMI 370
 Db 319 VAGDIPTGFMPOVEPRRLMQRVALLDAVALVAFAFSLMEMPARSHGYSVRANOELL 378
 Qy 371 ALGCSNFGSGFEKIHVICALSVTLAVDAGCKSOVASLCSVLYVMITMLVGLIYLP 430
 Db 379 AVGCCNVLPAFLHCPATSAALAKSLVKTATGCRTOSSVSATVLLVLLALAPLFHDQ 438
 Qy 431 KSVLCALLAVNLKNSLKQLTDPYIYMRKSKLDCCIMVVSFLSPFLPYGAVAVASV 490
 Db 439 RSVLACVIVSVLRGALRRKMGFPRLMRSPADALWAGTAAACMLVSTEGAGLAVILSL 498
 Qy 491 LVVVFQTFGRNGYALAOVMDTIYVNPNTYNRADIDQIKIITYCSPLYFANSEIFROK 550
 Db 499 LSLAGRTQRPRTALLARKIGDTAFYEDATEFEBLVEPEGVRYRRFGPLXYANKORFLOSL 558
 Qy 551 IAKTVSLOELQDFENAPPTDPNNNOTPANGTSVSYITFSPDSSPAOSEPPASAEADGE 610
 Db 559 ----YSLGILDAGCMARRKEGSETEGEG-----GPAQGEDLGP 595
 Qy 611 PSDMLASVPPPTFTHTLLIDMSGVSFVDMGLKALAKISTYTGKIGVAVFLNHAQYVN 670
 Db 596 VSTRALVPAAAGFTHTVVIDCAPLFLDLAAGVSTQDRLRDGALGISLILLACSSPPVAD 655
 Qy 671 DISHGVEFDG---SLECKHVPSTHDAVLEFAQANARPY--TPGH 710
 Db 656 ILSRGLFGEGRGDPDAEEQFLSVHDVAQVTRARHRELANTDAH 700

RESULT 14

Qy 09VVM6 PRELIMINARY: PRT; 742 AA.
 AC 09VVM6;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CG5485 PROTEIN.
 GN CG5485.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 CX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M.R., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt J., Nelson C.R., Miklos G.L.G.,
 Abell J.F., Asguyan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Boltsakov S.,
 Botkova D., Botchan M.A., Bouck J., Brokstein P., Brotlier P.,
 Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

Fri Apr 26 09:41:00 2002

us-09-749-589-2.rspt

Page 10

```

OY 138 ICDLAPESKQVONNMKPTNSESVDYATAMEABERLHVASTLACLTALATIONMGSGEPOEPAVL 197
Db 114 ARLRAPD--IPISNSNDINBSVPLGEEYVDPLFTSALLGLVGVOOIMGLIRLSEFLT 171
OY 198 YLSESFIRGEFTAAIGLOILSLVLEKYEIGLTIPSTYGEISVFEFIDICKMLPHTNIASLI 257
Db 172 YLSDLSVSGFTGAAYHVFTSOLKVKYGIKLPREGIGMIVRMVROIMSLGSVNRVALG 231
OY 258 FALISGAFLVLEKLMARIMKTI-----RPIPEMIVVVVATAISGGCKMPKKYMOY 312
Db 232 ISI----FGLIFDLGRYIIMPYIKKRSPIPPLELLIVFGIVISMIFMLDAEYHVKT 287
OY 313 GEIORGFTPVSPVSNQKMDIGAFASLIVSYINLAMRGLTANKHGVDSNOEMIL 372
Db 288 YEIRGFRPLSPRLNLPALLSLAIPAVVCTIWFVMSKGLPAKHKRYTDPQOELVAI 347
OY 373 GCSNFSGFFKIHVCCALSVTYLVADAGSKQVASCVLSVYMITVLVGLIYLPDPKS 432
Db 348 GIASLSSFFPVYVGVASLSSSVSCENSGANTQYITFSSFLITYILLGLPFLPELPMC 407
OY 433 VGLALIVANLKNLSKOLTDPYIYMRKSKLDCCIMVVSFSSFLSLPEYGAAGVARSVLV 492
Db 408 ILACIVIVSLKSLFMQVKEPLRLRISKYPAIMLVACSTIFTDVTGGLVSLASLYT 467
OY 493 VVQDFOFNGALVQWMDTIYVMPKRYNAODI-QGIKITTCSPLEFANSILFQKVI 551
Db 468 LVLRQQM--SFSTVLHDE--TPR-----QNPENWKVITFKAGSIHFENATAF----- 511
OY 552 AKTVSLOELQODFENAPPTDPNNQOPRANGTSVYITFSPDSSSPAOSBEPAPASAEAPGE 611
Db 512 -----QODM-----GEA 518
OY 612 SDMTASVPEVTEHTLLIDMSGVSFVLDIGIKALAKLSITYGIVKVELVNIHAOYND 671
Db 519 IGLPREDPRLDERTIILDDASSVAFIDIMGVDAIADVFKDALKLGVAHVFCGLPREDVL 578
OY 672 ISHGVEFEDGLECKHVPSIHDVLFQAONARV 706
Db 579 LSNDENFL-SVYPPSTFPST-DSCLLSFVHQSV 611

```

```
Search completed: April 26, 2002, 09:06:43
Job time: 134 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 09:04:39 ; Search time 12.77 Seconds

(without alignments)
2161.990 Million cell updates/sec

Title: us-09-749-589-2

Perfect score: 3869

Sequence: 1 MSQPRPRRYVDRAVSLTLE.....MDLEQMGSMFAETLTAL 753

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1190	30.8	780	1	PEND_HUMAN
2	1120	28.9	764	1	DRA_HUMAN
3	974.5	25.2	739	1	DTD_RAT
4	972	25.1	739	1	DTD_HUMAN
5	954.5	22.0	703	1	DTD_MOUSE
6	853	22.0	703	1	SAT1_RAT
7	588	15.2	758	1	YLDM_CAEL
8	518	13.4	662	1	SUT2_STYHA
9	516	13.3	667	1	SUT1_STYHA
10	496	12.8	644	1	SUT3_STYHA
11	464.5	12.0	877	1	SULH_SCHPO
12	437	11.3	754	1	SULX_YEAST
13	405	10.5	893	1	SULZ_YEAST
14	377	9.7	859	1	SULL_YEAST
15	375.5	9.7	485	1	NO70_SORBN
16	357	9.2	788	1	CY14_NEDCR
17	230.5	6.0	550	1	YCHM_ECOLI
18	181	4.7	434	1	YBAR_BACSU
19	135	3.5	1036	1	YG35_YEAST
20	132	3.4	442	1	CITN_LACIA
21	125	3.2	526	1	MYIN_TREPA
22	124	3.2	654	1	PSTA_MYGE
23	121	3.1	958	1	YA7B_SCHPO
24	121	3.1	1276	1	MR1_MOUSE
25	119.5	3.1	521	1	ND2C_SYNF7
26	116.5	3.0	445	1	YIEG_ECOLI
27	116	3.0	568	1	PTLB_STRMU
28	115	3.0	633	1	Y147_HAELN
29	114.5	3.0	400	1	NHAA_HAELN
30	113.5	2.9	470	1	YICO_ECOLI
31	113.5	2.9	533	1	MYIN_SYNF3
32	113.5	2.9	636	1	CTR4_HUMAN
33	113.5	2.9	1232	1	B3A3_HUMAN

34	112.5	2.9	471	1	NORM_CAUCR
35	111.5	2.9	525	1	YMT1_HUMAN
36	111.5	2.9	545	1	N05M_ALBCO
37	110.5	2.9	394	1	BERNE_ACICA
38	110.5	2.9	715	1	LCNC_LACIA
39	110.5	2.9	763	1	RGF2_YEAST
40	109.5	2.8	434	1	YHFT_ECOLI
41	109.5	2.8	744	1	N05C_GERJA
42	109	2.8	452	1	NORM_BACSU
43	108.5	2.8	438	1	SECY_METVA
44	108	2.8	561	1	B105_YEAST
45	107.5	2.8	699	1	N05C_DIGGR

ALIGNMENTS

RESULT 1

ID	PEND_HUMAN	STANDARD:	PRT:	780 AA.
AC	043511; 043170;			
DT	15-JUL-1999 (Rel. 38, Created)			
DT	15-JUL-1999 (Rel. 38, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	PENDRIN (SODIUM-INDEPENDENT CHLORIDE/IODIDE TRANSPORTER).			
GN	PDS.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND VARIANT PDS CYS-667.			
RC	TISSUE=Thyroid;			
RX	MEDLINE=96061089; PubMed=9398842;			
RA	Everett L.A., Glaser B., Beck J.C., Idol J.R., Buchs A., Heyman M.,			
RA	Adavi F., Hazani E., Nassir E., Baxevanis A.D., Sheffield V.C.,			
RA	Green E.D.;			
RT	"Pendred syndrome is caused by mutations in a putative sulphate			
RT	transporter gene (PDS).";			
RL	Nat. Genet. 17:411-422(1997).			
RN	[2]			
RP	SEQUENCE OF 336-780 FROM N.A.			
RA	Smith A., Johnson D., Harmon G.;			
RL	Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	FUNCTION.			
RX	MEDLINE=99206617; PubMed=10192399;			
RA	Scott D.A., Wang R., Kreman T.M., Sheffield V.C., Karishki L.P.;			
RT	"The Pendred syndrome gene encodes a chloride-iodide transport			
RT	protein.";			
RL	Nat. Genet. 21:440-443(1999).			
RN	[4]			
RP	VARIANTS PDS PRO-236 AND PRO-416.			
RX	MEDLINE=98282235; PubMed=9618166;			
RA	van Haave P., Everett L.A., Coucke P., Scott D.A., Kraft M.L.,			
RA	Ris-Stalpers C., Bolder C., Otten B., de Vilder J.J.M.,			
RA	Dierich N.L., Ramesh A., Srisailapathy S.C.R., Parving A.,			
RA	Cremers C.W.R.J., Millems P.J., Smith R.J.H., Green E.D., van Camp G.;			
RT	"Two frequent missense mutations in Pendred syndrome.";			
RL	Hum. Mol. Genet. 7:1099-1104(1998).			
RN	[5]			
RP	VARIANTS PDS PRO-236; GLY-384 AND PRO-416.			
RX	MEDLINE=98282236; PubMed=9618167;			
RA	Coffey B., Reardon W., Herbrick J.A., Tsui L.-C., Gausden E., Lee J.,			
RA	Coffey R., Grelleters A., Grossman A., Phelps P.D., Luxon L.,			
RA	Kendall-Taylor P., Scherer S.W., Trembath R.C.;			
RT	"Molecular analysis of the PDS gene in Pendred syndrome (sensorineural			
RT	hearing loss and goitre).";			
RL	Hum. Mol. Genet. 7:1105-1112(1998).			
RN	[6]			
RP	VARIANT DEFBA SER-497.			
RX	MEDLINE=98160177; PubMed=9500541;			
RA	Li X.C., Everett L.A., Lalwani A.K., Desmukh D., Friedman T.B.,			

FT	TRASMEM	653	673	11 (POTENTIAL).
FT	DOKAIN	674	780	EXTRACELLULAR (POTENTIAL).
FT	CAROHYD	415	415	O-LINKED (POTENTIAL).
FT	VARIANT	209	209	G -> V (IN EVA).
FT	VARIANT	236	236	/FTID-VAR..007440.
FT	VARIANT	236	236	L -> P (IN PDS).
FT	VARIANT	369	369	/FTID-VAR..007441.
FT	VARIANT	372	372	K -> E (IN EVA).
FT	VARIANT	372	372	/FTID-VAR..007442.
FT	VARIANT	384	384	A -> V (IN EVA).
FT	VARIANT	384	384	/FTID-VAR..007443.
FT	VARIANT	416	416	E -> G (IN PDS).
FT	VARIANT	416	416	/FTID-VAR..007444.
FT	VARIANT	497	497	T -> P (IN PDS).
FT	VARIANT	497	497	/FTID-VAR..007445.
FT	VARIANT	667	667	G -> S (IN DENB4).
FT	VARIANT	667	667	/FTID-VAR..007446.
FT	VARIANT	721	721	F -> C (IN PDS).
FT	VARIANT	723	723	/FTID-VAR..007447.
FT	VARIANT	723	723	T -> M (IN EVA).
FT	VARIANT	723	723	/FTID-VAR..007448.
FT	VARIANT	723	723	H -> R (IN EVA).
FT	VARIANT	723	723	/FTID-VAR..007449.
FT	VARIANT	723	723	L -> M (IN REF..2).
FT	VARIANT	723	723	/FTID-VAR..007450.
FT	VARIANT	723	723	/FTID-VAR..007451.
FT	VARIANT	723	723	/FTID-VAR..007452.
FT	VARIANT	723	723	/FTID-VAR..007453.
FT	VARIANT	723	723	/FTID-VAR..007454.
FT	VARIANT	723	723	/FTID-VAR..007455.
FT	VARIANT	723	723	/FTID-VAR..007456.
FT	VARIANT	723	723	/FTID-VAR..007457.
FT	VARIANT	723	723	/FTID-VAR..007458.
FT	VARIANT	723	723	/FTID-VAR..007459.
FT	VARIANT	723	723	/FTID-VAR..007460.
FT	VARIANT	723	723	/FTID-VAR..007461.
FT	VARIANT	723	723	/FTID-VAR..007462.
FT	VARIANT	723	723	/FTID-VAR..007463.
FT	VARIANT	723	723	/FTID-VAR..007464.
FT	VARIANT	723	723	/FTID-VAR..007465.
FT	VARIANT	723	723	/FTID-VAR..007466.
FT	VARIANT	723	723	/FTID-VAR..007467.
FT	VARIANT	723	723	/FTID-VAR..007468.
FT	VARIANT	723	723	/FTID-VAR..007469.
FT	VARIANT	723	723	/FTID-VAR..007470.
FT	VARIANT	723	723	/FTID-VAR..007471.
FT	VARIANT	723	723	/FTID-VAR..007472.
FT	VARIANT	723	723	/FTID-VAR..007473.
FT	VARIANT	723	723	/FTID-VAR..007474.
FT	VARIANT	723	723	/FTID-VAR..007475.
FT	VARIANT	723	723	/FTID-VAR..007476.
FT	VARIANT	723	723	/FTID-VAR..007477.
FT	VARIANT	723	723	/FTID-VAR..007478.
FT	VARIANT	723	723	/FTID-VAR..007479.
FT	VARIANT	723	723	/FTID-VAR..007480.
FT	VARIANT	723	723	/FTID-VAR..007481.
FT	VARIANT	723	723	/FTID-VAR..007482.
FT	VARIANT	723	723	/FTID-VAR..007483.
FT	VARIANT	723	723	/FTID-VAR..007484.
FT	VARIANT	723	723	/FTID-VAR..007485.
FT	VARIANT	723	723	/FTID-VAR..007486.
FT	VARIANT	723	723	/FTID-VAR..007487.
FT	VARIANT	723	723	/FTID-VAR..007488.
FT	VARIANT	723	723	/FTID-VAR..007489.
FT	VARIANT	723	723	/FTID-VAR..007490.
FT	VARIANT	723	723	/FTID-VAR..007491.
FT	VARIANT	723	723	/FTID-VAR..007492.
FT	VARIANT	723	723	/FTID-VAR..007493.
FT	VARIANT	723	723	/FTID-VAR..007494.
FT	VARIANT	723	723	/FTID-VAR..007495.
FT	VARIANT	723	723	/FTID-VAR..007496.
FT	VARIANT	723	723	/FTID-VAR..007497.
FT	VARIANT	723	723	/FTID-VAR..007498.
FT	VARIANT	723	723	/FTID-VAR..007499.
FT	VARIANT	723	723	/FTID-VAR..007500.
FT	VARIANT	723	723	/FTID-VAR..007501.
FT	VARIANT	723	723	/FTID-VAR..007502.
FT	VARIANT	723	723	/FTID-VAR..007503.
FT	VARIANT	723	723	/FTID-VAR..007504.
FT	VARIANT	723	723	/FTID-VAR..007505.
FT	VARIANT	723	723	/FTID-VAR..007506.
FT	VARIANT	723	723	/FTID-VAR..007507.
FT	VARIANT	723	723	/FTID-VAR..007508.
FT	VARIANT	723	723	/FTID-VAR..007509.
FT	VARIANT	723	723	/FTID-VAR..007510.
FT	VARIANT	723	723	/FTID-VAR..007511.

Db 613 NAFPEDEDIEDLEEDLDPTKTEIELQVOWNSBLPRKVVNPKVPHISLVLDGALSFLDVG 672Z

QY 642 IKALAKLSSTFYKIGCVKFLVNIHAQYYINDSHGCVFEDGSLECKHVPFPIHVAFLAQ 700

673 VRSLSRVYKEQRDLVAVNVFASLDQDYIEKLEQCGFFDD-NIRKDTFFLVYHDAILVLQ 730

```

RESULT 2
DRA_HUMAN
ID DRA_HUMAN STANDARD; PRT; 764 AA.
AC P40879;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DRA PROTEIN (DOWN-REGULATED IN ADENOMA).
GN DRA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-COLON:
RX MEDLINE-93248250; PubMed-7683425;
RA Schweinfest C.W., Henderson K.W., Sustar S., Kondoh N., Papas T.S.;
RT "Identification of a colon mucosa gene that is down-regulated in
RL colon adenomas and adenocarcinomas.";
RN Proc. Natl. Acad. Sci. U.S.A. 90:4166-4170(1993).
RN [2]
RP SIMILARITY TO SULFATE PERMEASES.
RX MEDLINE-94188926; PubMed-8140616;
RA Sandal N.N., Marcker K.A.;
RT "Similarities between a soybean nodulin, Neurospora crassa sulphate
RL permease II and a putative human tumour suppressor.";
RN Trends Biochem. Sci. 19:19-19(1994).
RN [3]
RP VARIANTS CTD LEU-124 AND VAL-317 DEL. AND VARIANT TRP-307.
RX MEDLINE-97051927; PubMed-8896562;
RA Hoeglund P., Halla S., Socha J., Tomaszewski L., Saarialho-Kere U.,
RA Karjalainen-Lindberg M.-L., Altiola K., Holmberg C.,
RA de la Chapelle A., Kere J.;
RT "Mutations of the down-regulated in adenoma (DRA) gene cause
RL congenital chloride diarrhoea.";
RN Nat. Genet. 14:316-319(1996).
RN [4]
RP VARIANTS CTD SER-120; ARG-131; VAL-317 DEL AND TYR-527 DEL.
RX MEDLINE-98213471; PubMed-9554749;
RA Hoeglund P., Halla S., Gustavson K.-H., Taipale M., Hannula K.,
RA Poplonska K., Holmberg C., Socha J., de la Chapelle A., Kere J.;
RT "Clustering of private mutations in the congenital chloride
RL diarrhoea/down-regulated in adenoma gene.";
RN Hum. Mutat. 11:321-327(1996).
RN [5]
RP FUNCTION: POSSIBLE SULFATE TRANSPORTER.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- DEVELOPMENTAL STAGE: EXPRESSION IS SIGNIFICANTLY DECREASED IN
CC ADENOMAS (POLYPS) AND ADENOCARCINOMAS OF THE COLON.
CC -1- DISEASE: DEFECTS IN DRA ARE THE CAUSE OF CONGENITAL CHLORIDE
CC DIARRHEA (CLD), A DISEASE CHARACTERIZED BY VOLUMINOUS WATERY
CC STOOLS CONTAINING AN EXCESS OF CHLORIDE. THE CHILDREN WITH THIS
CC DISEASE ARE OFTEN PREMATURE.
CC -1- SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, L02785; AAAS8443.1; -.
CC MIM, 126550; -.
CC MIM, 214700; -.
DR

```

DR	InterPro:IPR002645; STAS.
DR	InterPro:IPR001903; Sulfate_transp.
DR	Pfam: PF01740; STAS: 1.
DR	Pfam: PF00916; Sulfate_transp: 1.
DR	PROSITE; PS01130; SULFATE_TRANSP: 1.
KM	Transposite; Transmembrane; Disease mutation; Polymorphism.
FT	TRANSPOSIT 77 97 POTENTIAL.
FT	TRANSMEM 100 120 POTENTIAL.
FT	TRANSMEM 125 145 POTENTIAL.
FT	TRANSMEM 176 196 POTENTIAL.
FT	TRANSMEM 198 218 POTENTIAL.
FT	TRANSMEM 259 279 POTENTIAL.
FT	TRANSMEM 286 306 POTENTIAL.
FT	TRANSMEM 343 363 POTENTIAL.
FT	TRANSMEM 375 395 POTENTIAL.
FT	TRANSMEM 412 432 POTENTIAL.
FT	TRANSMEM 439 459 POTENTIAL.
FT	TRANSMEM 470 490 POTENTIAL.
FT	TRANSMEM 644 664 POTENTIAL.
FT	TRANSMEM 702 722 POTENTIAL.
FT	VARIANT 120 120 POTENTIAL.
FT	VARIANT 124 124 G -> S (IN CLD).
FT	VARIANT 131 131 H -> L (IN CLD).
FT	VARIANT 131 131 /FtId=VAR.007429.
FT	VARIANT 131 131 P -> R (IN CLD).
FT	VARIANT 131 131 /FtId=VAR.007430.
FT	VARIANT 307 307 C -> W (PROBABLE POLYMORPHISM).
FT	VARIANT 317 317 /FtId=VAR.007431.
FT	VARIANT 317 317 MISSING (IN CLD).
FT	VARIANT 527 527 /FtId=VAR.007432.
FT	VARIANT 527 527 MISSING (IN CLD).
FT	VARIANT 527 527 /FtId=VAR.007433.
FT	SEQUENCE 764 AA; 84504 MW; 6594CSBC2D4121F6D CRC64;

	Query Match	28.9%	Score 1120:	DB 1:	Length 764:
	Best Local Similarity	34.0%	Pred. No. 2,46-67:		
	Matches 250:	Conservative 142:	Mismatches 275:	Indels 68:	Gaps 11:
QY	7	RYVVDRAAYSLTLDDEFERKDRTPVG-EKLRNAPRCCSSAKIKAVNGLLPVLWSLPKY	65		
	8	QYIARAPVYSTNAEEENKKTGTGRHNRKTPDLHDKVCCSCSPQAKRIVLSLPIASMLPAY	67		
QY	66	KIKDYIIPDLGGSGSSIOYPOGAAFPALNLPVNGICGSSSEFPDLLYFFLFGVHONWP	125		
Db	68	RKLEMLSDIYSGISGTIVAVLQGLAEFLALVDIPRYVYGLIYASFPAITLYLFFGTSRHNISV	127		
QY	126	GTFEAVISILVG---NICLOLAPESKPOVF--NNATNESYVDTAAMEAERLHVSATLAC	178		
Db	128	GPFFILSMWVGLAVSAGVASKAVPDRNATITLGLPNSSNNSSILDD--ERVRAAASAATV	184		
QY	179	LTAITOMGLGFMQGFVAIYIYSESEIRGEFMPAAGIOLILSVKTIFFGLTIPSTYGPSSIV	238		
Db	185	LSGIIOALFGLIRGEFVYIYLSSELSISGFTTAAAHVILVYSLQKLFQOLLPSHTDPVSTF	244		
QY	239	FTEIDICKNLPHNTIASLIFALISGAPFLVYKELNARVYHNRKIRPDIPTMIYVVAATAIS	298		
Db	245	KYLIVSEQIOIEKTNLADLVLTALIVLVVSIYKEINORFRDKLPVPIPIEFIMTVIAAGVS	304		
QY	299	GGCKMPKRYHMOIYGEIQRGFPVPVPSVQMKMDIGAFSLAISVYINLAMGRTLANK	358		
Db	305	YGCDEKRNFRKAVVAGDMNMGFQPPITTPVETFEQNTVGCDFGIANVAFAAVAFSVASVSLK	364		
QY	359	HGYDVSQDQMIALGCSNFFGSFKEIYHICCLSVLTAVDAGGKGSQVASLCSVLVYMT	418		
Db	365	YDYPIDGNOELIATLGNIVCCVFGEFGRGASTLASVAOESGKGTOIAGLLGAIITIVLV	424		
QY	419	MLVLIGIYLPPLKSVYGLALIANVLEKRSIKQLDPYVLRKSKLDDCCIVVSELSSEFSL	478		
Db	425	VLAIGFLIAPLQKSVLALALGNLKGMLMQFIEIRLWRMKDKYDDCLIMTFELFTIVIGL	484		
QY	479	PYGAVAAGVAFSVLVVYEQTFQFRNGYALAAQVMDTIYVANKPTYNRRADIODIGIITITCSP	538		

```

Db 485 GIGLASAVAFOLLITVFRTQFPCSTLANIGRTNKKKDYDYMEPEGVKIFRCPSP1 544
Oy 539 YFANSEIFROKYI-----AKTYSLOELOQ-----DFENAP 568
| | | | | : | | | | |
Db 545 YFANIGFPRKILDAVGFSPRLRKRNKALKRIRKLOGLQVTPKFCITGVDIKXS 604
Oy 569 PFDPRNNQ-----TPANGISVST-ITFSPDSSSPASEPPASAEAGCEPSDMLASVPPV 622
| | | | | : | | | | |
Db 605 DEELDNNOIEVLDOPIINTDLPFHIDWMDL-----PLNIEVFK-----I 644
Oy 623 TFFHTLLDMSGVFDVLMGKALKAKLSYTGKIGVKEFLVNIHAQVYNDISHGVEEDGS 682
| | | | | : | | | | |
Db 645 SLHSLLDPSAVSFVDVSSVYRGKSLIOEFTIKKYDVIYIGDDDFTEKLNIEFF-DGE 703
Oy 683 LECKHVFPSIDAVL 697
| | | | | : | | | | |
Db 704 VKSIEFFLTIHDAVL 718

RESULT 3
DID_RAT STD_RAT STANDARD: PRT: 739 AA.
AC 070531:
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SUIPATE TRANSPORTER (DIASTROPHIC DYSPLASIA PROTEIN HOMOLOG).
GN SLC26A2 OR DTDST.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN 11)
RP TSSSEQUENCE FROM N.A.
RC TISSUE=Osteosarcoma;
RX MEDLINE=98241598; PubMed=9575183;
RA Satoh H., Suuaki M., Shukunami C., Iyama K., Negoro T., Hiraki Y.;
RT "Functional analysis of diastrophic dysplasia sulfate transporter.
RT Its involvement in growth regulation of chondrocytes mediated by
RT sulfated proteoglycans."
RL J. Biol. Chem. 273:12307-12315(1998).
CC -1- FUNCTION: SUIPATE TRANSPORTER. MAY PLAY A ROLE IN ENDOCHONDRAL
CC BONE FORMATION.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: CARTILAGE AND INTESTINE.
CC -1- SIMILARITY: BELONGS TO THE SULFATE PEROXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, D82883: BA25987.1; -
DR InterPro: IPR002645: STAS.
DR InterPro: IPR001902: Sulfate_transp.
DR Pfam: PF01740: STAS; 1.
DR Pfam: PF00916: Sulfate_transp. 1.
DR PROSITE: PS01130: SULFATE_TRANS. 1.
KW Transport; Transmembrane; Glycoprotein.
FT DOMAIN 1 111 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 112 132 POTENTIAL.
FT DOMAIN 133 136 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 137 157 POTENTIAL.
FT DOMAIN 158 159 POTENTIAL.
FT TRANSMEM 160 180 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 181 218 POTENTIAL.
FT TRANSMEM 219 239 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 240 241 POTENTIAL.
FT TRANSMEM 242 262 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 263 296 POTENTIAL.
FT EXTRACELLULAR (POTENTIAL).

```

```

FT TRANSMEM 297 317 POTENTIAL.
FT DOMAIN 318 328 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 329 349 POTENTIAL.
FT DOMAIN 350 377 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 378 398 POTENTIAL.
FT DOMAIN 399 420 POTENTIAL.
FT TRANSMEM 421 441 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 442 454 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 455 475 POTENTIAL.
FT DOMAIN 476 523 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 524 544 POTENTIAL.
FT DOMAIN 545 643 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 644 664 POTENTIAL.
FT DOMAIN 665 739 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 205 205 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 357 357 N-LINKED (GLCNAc. . .) (POTENTIAL).
SQ SEQUENCE 739 AA: 82027 MW: 0C539F66D478C8DA CXC64;

Query Match 25.2%; Score 974.5; DB 1: Length 739;
Best Local Similarity 31.7%; Pred. No. 1.2e-57;
Matches 219; Conservative 140; Mismatches 280; Indels 51; Gaps 8;

Oy 33 VGEKLNARFSSAKIKAVFGLPLVSLPKYTKNDYIIPDLGGLSGSGSIQVPOGMAR 92
| | | | | : | | | | |
Db 70 VVRKLOKSCQCNATKIRINRIFDFPFVLRWLPYDKRKNIGDMGSLIVGILLVPOSIAY 129
Oy 93 ALLANIPAVNGYSSFFPLTFYFLGVHQPVGTFEAVSILVGN-----CLOL-A 143
| | | | | : | | | | |
Db 130 SILAQQEPIYGLTYTFEFAIYIFLEGTSHRISVIGFGLICLMIGEVDRELHKACPDIDT 189
Oy 144 PESKFOVENN--ATNESYVDTAAMEABERLHVSATLACTALITIQMGLFMQGFVAIYLS 200
| | | | | : | | | | |
Db 190 TSSSIAMFSGNCVVNHTLDGLCDKSCYAIRKIGSTVFAGVAYQAMGFQGVFVSVLS 249
Oy 201 EEFINGFMAAGLOLLISLKYIEGLTIPSYGPGSIYFTFDICKNLRHNTAALIRAL 260
| | | | | : | | | | |
Db 250 DALLSGFVYTGASFTLLTSQAKYLLSLPRNSGVSVITWTHIRNIRKNIKCOLITSL 309
Oy 261 ISGAFVLVLEKLNARVMHRIKPIPTEMIVVVAIAISGCGMPKRYHQVIGEIORGFP 320
| | | | | : | | | | |
Db 310 LCLLVLPYRKELNEFKSKLAPRIPTELIVVAATLASHFGKLNINYSIAGQIPTGFM 369
Oy 321 TPVSPVVSQWKMDIGTAFSLAIVSYVNIAMGRITLANKRGYDVDSNOEMIALGCSNFGS 380
| | | | | : | | | | |
Db 370 PQQAPDWSLIPVAVDAIAISIIIGFAITVSLSEMPAKKHGYVKKANQEMVAIGFCNIIPS 429
Oy 381 PFKIHVICALSVTLANVGAGKSOVASLCSLVYIMTLVIGITLYPLRPSVIGALLAY 440
| | | | | : | | | | |
Db 430 PFHCITTSAAALAKTLVKESTGQOTLSAIVSLVLLVLLAPLFLYSLOKCVLAVITIV 489
Oy 441 NLKNSLKOLTDYVLMRKSKLDCCIMVVSFLSPFLSPYGVAVGVAVSVLVVPOFOR 500
| | | | | : | | | | |
Db 490 NLRGALLFRDLPKMKRSLRMDTYIMFTVMSALLSTREIGLIVGVCSCMVCVILRTOMP 549
Oy 501 NGYALAQVMDTDIVNPKTYNRAODIGIKITTYCSPLYFANSEIFROKYIAKTYSLOEL 560
| | | | | : | | | | |
Db 550 KISLIGLEESSEIFESISTYKMLRSKSGIKVERFAPLTYINKKECFKSAIYKKTLPNPLV 609
Oy 561 QODEFNAPTDPPNNQNTANGTSVSYITFSPDSSSPASEPPASAEAGCEPSDMLASVPP 620
| | | | | : | | | | |
Db 610 KAAMKKAARKRLKEET-----VFHGD-----PDEVSMQLSHDP- 643
Oy 621 FVTFTLLDMSGVFDVLMGKALKAKLSYTGKIGVKEFLVNIHAQVYNDISHGVED 680
| | | | | : | | | | |
Db 644 -LEHTVYIDCSAIOFLDTAGIHITLKEVRRDYEALIGIYVLLAQCNPSVDSLANGEY--- 699
Oy 681 GSLECKH-----VPSIDAVLFAQANARD 705
| | | | | : | | | | |
Db 700 ---CKKREENILFYSLSAEAVAFAEESQKE 725

RESULT 4

```

ID	DTD_HUMAN	STANDARD;	PT;	739 AA.
AC	P50443:			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	SULFATE TRANSPORTER (DIASTROPHIC DYSPLASIA PROTEIN).			
GN	SIC26A2 OR DTDSPT OR DTD.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_Taxid=9606;			
RT	[1]			
RL	SEQUENCE FROM N.A.			
RP	MEDLINE=95007757; PubMed=7923357;			
RX	Haestbacka J., de la Chapelle A., Mahtani M.M., Clines G.,			
RA	Reeve-Daly M.P., Daly M., Hamilton B.A., Kusumi K., Trivedi B.,			
RA	Weaver A., Coloma A., Lovett M., Buckler A., Kaitila I., Lander E.S.;			
RT	"The diastrophic dysplasia gene encodes a novel sulfate transporter:			
RL	positional cloning by fine-structure linkage disequilibrium			
RL	mapping."			
RL	Cell 78:1073-1087(1994).			
RP	[2]			
RP	VARIANTS ACG-1B VAL-340 DEL; ASP-425 AND VAL-678.			
RX	MEDLINE=96122050; PubMed=8528239.			
RA	Superti-Furga A., Haestbacka J., Wilcox W.R., Cohn D.H.,			
RA	van der Harten H.J., Rossi A., Blau N., Rimoin D.L., Steinmann B.,			
RA	Lander E.S., Gitzeleimann R.;			
RT	"achondrogenesis type IB is caused by mutations in the diastrophic			
RT	dysplasia sulphate transporter gene."			
RL	Nat. Genet. 12:100-102(1996).			
RP	[3]			
RP	VARIANTS AO-II GLU-255; TRP-279 AND VAL-715.			
RX	MEDLINE=96152121; PubMed=8571951;			
RA	Haestbacka J., Superti-Furga A., Wilcox W.R., Rimoin D.L., Cohn D.H.,			
RA	Lander E.S.;			
RT	"Aletostegenesis type II is caused by mutations in the diastrophic			
RT	dysplasia sulphate-transporter gene (DTDSPT): evidence for a phenotypic			
RT	series involving three chondrodysplasias."			
RL	Am. J. Hum. Genet. 58:255-262(1996).			
CC	-I- FUNCTION: SULFATE TRANSPORTER. MAY PLAY A ROLE IN ENDOCHONDRAL			
CC	BONE FORMATION.			
CC	-I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.			
CC	-I- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED.			
CC	-I- DISEASE: DEFECTS IN DTDSPT ARE THE CAUSE OF DIASTROPHIC DYSPLASIA			
CC	(DTD), AN AUTOSOMAL RECESSIVE DISEASE CHARACTERIZED BY			
CC	OSTEOCHONDRODYSPLASIA WITH CLINICAL FEATURES INCLUDING DWARFISM,			
CC	SPINAL DEFORMATION, AND SPECIFIC JOINT ABNORMALITIES.			
CC	-I- DISEASE: DEFECTS IN DTDSPT ARE THE CAUSE OF ACHONDROGENESIS TYPE IB			
CC	(ACG-1B). ACG-1B IS A RECESSIVELY INHERITED CHONDRODYSPLASIA			
CC	CHARACTERIZED BY EXTREMELY POOR SKELETAL DEVELOPMENT AND PERINATAL			
CC	DEATH.			
CC	-I- DISEASE: DEFECTS IN DTDSPT ARE THE CAUSE OF ALESTOGENESIS TYPE			
CC	II (AO-II), ALSO KNOWN AS NEONATAL OSSOUS DYSPLASIA I, WHICH IS			
CC	CHARACTERIZED BY SEVERELY SHORTENED LIMBS, SMALL CHEST, SCOLIOSIS,			
CC	CLUB FOOT OF THE EQUINOVARUS TYPE (PALPES EQUINOVARUS), ABDUCTED			
CC	THUMBS AND GREAT TOES, AND CLEFT PALATE. PATIENTS DIED OF			
CC	RESPIRATORY INSUFFICIENCY SHORTLY AFTER BIRTH BECAUSE OF THE			
CC	COLLAPSE OF THE AIRWAYS AND PULMONARY HYPOPLASIA DUE TO THE SMALL			
CC	RIB CAGE.			
CC	-I- SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL: U14528; AAA70081.1; -			
DR	MTM: 222600; -			
DR	MTM: 600972; -			

DR	MM: 256050; -	IPR002645; STAS.
DR	InterPro: IPR001902; Sulfate_transp.	
DR	Pfam: PF01740; STAS: 1.	
DR	Pfam: PF00916; Sulfate_transp: 1.	
DR	PROSITE: PS01130; SULFATE_TRANS: 1.	
KM	Transprot: Transmembrane; Glycoprotein; Disease mutation.	
FT	DOMAIN 1	111
FT	TRANSSEM 112	132
FT	DOMAIN 133	136
FT	TRANSSEM 137	157
FT	DOMAIN 158	159
FT	TRANSSEM 160	180
FT	DOMAIN 181	218
FT	TRANSSEM 219	239
FT	DOMAIN 240	241
FT	TRANSSEM 242	262
FT	DOMAIN 263	296
FT	TRANSSEM 297	317
FT	DOMAIN 318	328
FT	TRANSSEM 329	349
FT	DOMAIN 350	377
FT	TRANSSEM 378	398
FT	DOMAIN 399	420
FT	TRANSSEM 421	441
FT	DOMAIN 442	454
FT	TRANSSEM 455	475
FT	DOMAIN 476	523
FT	TRANSSEM 524	544
FT	DOMAIN 545	643
FT	TRANSSEM 644	664
FT	DOMAIN 665	739
FT	CARBOHYD 199	199
FT	CARBOHYD 205	205
FT	CARBOHYD 357	357
FT	VARIANT 255	255
FT	VARIANT 279	279
FT	VARIANT 340	340
FT	VARIANT 425	425
FT	VARIANT 678	678
FT	VARIANT 715	715
SO	SEQUENCE 739 AA; 81649 MW; BD613DCB650D549 CRC64;	
Query Match	25.1%; Score 972; DB 1; Length 739;	
Best Local Similarity	31.2%; Pred. NO. 1.7e-57;	
Matches	226; Conservative 138; Mismatches 285; Indels 76; Gaps 11.	
OY	3 QPRPVVVDRAVYSLTLPDDEFEKDRTP--VGEKLRNAPFCSSAKIRAAVFGLLPVLS 60	
DB	49 RPYHILLERQ-----EKSDTNEKEFVIRKLLQKNCOCSPAKAKNMILGELPVQ 97	
OY	61 WLPKKIKDIYIIPDLGLSGSGSIOVPGCMARFALLANLPVAVGLVSSFPPLITLYPFLGCV 120	
DB	98 WLPKIDELKNLIGDVGSGILVILVLPSSIASYSLAGLGEPEYGLVTSFPASITIFPLGTS 157	
OY	121 HQWVGTFRAVISILVGNIC--LQ-----LAPESKFOVFNATNESYVDTTAM 165	
DB	158 RHISVIGIGVCLMIGETVYDRELQAGYDQNAHSAPSLGWSGSTLLNHTSDRICKSCY 217	
OY	166 EAERLHVSATLACLTATITQMLGEMQFGVAIYLSBSFTRGMTAAGDILISVLKYITG 225	
DB	218 --AIMGSTVFYFIAGVGVAMGFQVGVSYLLSDALLSGCVTSGASFTILTSQAAYLLG 274	
OY	226 LTIPTVPGSVTFEFDIDCKNLKPLPTNIALSLFALISGAFILVVELNARYMHKIRFP 285	
DB	275 LNLPTNGVSGITTTMHVFNIRHNTCDLLTSLCLLVLLPTELELNHFSKILAPR 334	

[illegible]

```

RESULT 5
DTD_MOUSE
ID DTD_MOUSE STANDARD: PRT: 739 AA.
AC Q6273:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SULFATE TRANSPORTER (DIASTROPHIC DYSPLASIA PROTEIN HOMOLOG) (ST-OB).
GN SLC26A2 OR DTDST OR DTD.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE:98036067; Pubmed-9370300;
RA Kobayashi T., Sugimoto T., Saijoh K., Fukase M., Chihara K.;
RT Cloning of mouse diastrophic dysplasia sulfate transporter gene
RT induced during osteoblast differentiation by bone morphogenetic
RT protein-2."
RL Gene 198:341-349(1997).
CC -1- FUNCTION: SULFATE TRANSPORTER. MAY PLAY A ROLE IN ENDOCHONDRAL
CC BONE FORMATION.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: DISTRIBUTED MAINLY IN THE THYMUS, TESTIS AND
CC OSTEOBLASTIC CELLS.
CC -1- SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC -----
DR EMBL, D42049; BAA07650.1; -.
DR MGD; MGI:892977; SLC26a2.
DR InterPro: IPR002645; STR
DR

```

DR		Interfco: IPRO001902; Sulfate_transp.
DR	Pfam: PF017740; STAS: 1.	Sulfate_transp. 1.
DR	Pfam: PF00916; Sulfate_transp. 1.	Sulfate_transp. 1.
DR	ProSite: PS01130; SULFATE_TRANS_1.	Glycoprotein.
KV	TransPort: Transmembrane; Glycoprotein.	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 1	111
FT	TRANSMEM	112 132
FT	DOMAIN	133 136
FT	TRANSMEM	137 157
FT	DOMAIN	138 159
FT	TRANSMEM	160 180
FT	DOMAIN	181 218
FT	TRANSMEM	219 239
FT	DOMAIN	240 241
FT	TRANSMEM	242 262
FT	DOMAIN	263 296
FT	TRANSMEM	297 317
FT	DOMAIN	318 328
FT	TRANSMEM	329 349
FT	DOMAIN	350 377
FT	TRANSMEM	378 398
FT	DOMAIN	399 420
FT	TRANSMEM	421 441
FT	DOMAIN	442 454
FT	TRANSMEM	455 475
FT	DOMAIN	476 523
FT	TRANSMEM	524 544
FT	DOMAIN	545 643
FT	TRANSMEM	644 664
FT	DOMAIN	665 739
FT	CARBOHYD	205
FT	CARBOHYD	357
SO	SEQUENCE	739 AA; 81604 MW; EEDID6B278ECBD CRG64;

[illegible]


```

Db 550 KNSLGLSESEFESISTYKNIKRSKGIVKVFIAPIPLYINKECKRSALYKALNPVLV 609
Qy 561 QODFENAPPTDNNQTPANGTSVSYTFSPDSSPPAOSPEPPASABAPGEPDMLASVPP 620
Db 610 KAMKKAARKRIKLEEM-----VTFRGD-----PDEVSNQLSHDP- 643
Qy 621 FVTHTLLIDMGSGVSPVDLMGKALAKLSITGKIGKVFVNIHAQYINDISHGVEED 660
Db 644 -LEVHTIVDCSAIOFLDPAGIHTTLEKVERDYEAQVIOVLLAQCNPSVRSLARGEY--- 699
Qy 681 GSLECKH-----VFPSIHDAVLEAFO 700
Db 700 ----CKKEETLTFTSLSEAVAFAE 720

RESULT 6
SAT1_RAT
ID SAT1_RAT STANDARD: PRT: 703 AA.
AC P45380;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SULFATE ANION TRANSPORTER 1 (CANALICULAR SULFATE TRANSPORTER)
DE (SULFATE/CARBONATE ANTI-PORTER).
CN SAT-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RA MEDLINE=94132077; PubMed=8300633;
RA Bissig M., Hagenbuch B., Stieger B., Koller T., Meier P.J.;
RT "Functional expression cloning of the canalicular sulfate transport
RT system of rat hepatocytes."
RL J. Biol. Chem. 269:3017-3021(1994).
CC - FUNCTION: HIGH AFFINITY UPTAKE OF SULFATE. ACCEPTS OXALATE, BUT
CC NOT SUCCINATE AS A CO-SUBSTRATE.
CC - TISSUE CELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC - TISSUE SPECIFICITY: LIVER, KIDNEY. LESS ABUNDANT IN MUSCLE AND
CC BRAIN.
CC - SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L23413; AAA17545.1; -
DR InterPro: IPR002645; STAS.
DR InterPro: IPR001902; Sulfate_transp.
DR Pfam: PF01740; STAS; 1.
DR PROSITE: PS01130; SULFATE_TRANSP; 1.
KM Transmembrane; Glycoprotein.
FT TRANSMEM 76 96 POTENTIAL.
FT TRANSMEM 98 118 POTENTIAL.
FT TRANSMEM 187 207 POTENTIAL.
FT TRANSMEM 209 229 POTENTIAL.
FT TRANSMEM 260 280 POTENTIAL.
FT TRANSMEM 294 314 POTENTIAL.
FT TRANSMEM 341 361 POTENTIAL.
FT TRANSMEM 381 401 POTENTIAL.
FT TRANSMEM 418 438 POTENTIAL.
FT TRANSMEM 486 506 POTENTIAL.
FT TRANSMEM 606 626 POTENTIAL.
FT CARBOHYD 158 158 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 587 587 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

SQ SEQUENCE 703 AA: 75447 MW: A90D839D70BD2738 CRC64;
Query Match 22.0%; Score 853; DB 1; Length 703;
Best local similarity 29.3%; Pred. No. 1,4e-49;
Matches 201; Conservative 143; Mismatches 304; Indels 38; Gaps 8;
Qy 36 KLRNAPRCSSAKIRAVFGLLPVLSQLPKYKIDYIIPDLLGSLGSIQVPGQMAFALL 95
Db 34 RLKSCSTSPMCAQALVQGLFPIRMLPQYRLKEYLAGDVMSGLIGIILVPOAIAYSL 93
Qy 96 ANPVAINGLYSSFFPLTYFFLFGVHQMVRGTFAVISILVGNIC-----LQIA----- 143
Db 94 AGLOPITYSLYTSFFANLTYFLMGTSRHVNVGIFSLCLMGVQVVDRELQLAGPDSDSL 153
Qy 144 -PESKFOVFENN-ATNESYVDYTAAMEAEERLHVSAITLACTFAITIQMGJFGFVATYLS 201
Db 154 GPGNNSTLNTNTATLVYGLQDCGRDCHAIATATLTMAGLYQVLMGILRGVSTYLSQ 213
Qy 202 SFIRGEVTAAGLIQILISLVKIFGLTIPSYTGPGSIVTFPIDICKNLPHTNIASLIFALI 261
Db 214 PLIDFGAMGASVYILTSQAKHLGLVRIPRHQGLGMVHTMLSLQNVGQANLGDVYTSAV 273
Qy 262 SGAFVLVYKRLNRYMKIKFPIPTEMIVVYVATISGGCKMKRYMOMQVGEIQRGFT 321
Db 274 CLAVLLTAKELSDRYRYLYKVPVPTLLVIVATIASHFQGLHTRFGSSVAGNIPGFEVA 333
Qy 322 PPSPVVSQMKMDIGTAFSLAIVSVINLAMGRTLANKGYDVDSNOBMIALGCSNFGSF 381
Db 334 POIPDKRIMSVALLDMSLVLGVSAPFISLAEMFARSHGTSVSNQDELLAVGCCNVLP 393
Qy 382 FKIHVICALSVTLAVDAGSGQSVASLVVMITMLVIGLYLPKPSVGLALAVN 441
Db 394 FHCFAVSAAALSKTLVYKATCGQFOLSSVSAVAVLLVLIAPLPHDQICVLACTIVVS 453
Qy 442 LKNSKQLTDPYTLMKRSKIDCCIVWSFLSPFLSPYGVANGVAVSVLVVYVQYQFRN 501
Db 454 LRGLARVKDLPQMLRISPDALVWATATCVLVSIEADLLAGVFSLSLAGRQRP 513
Qy 502 GYVLAQVMDIDIVNPPTRYRAODIGIKITCYSPFYFNSEIFRQKVIATKYSIQELQ 561
Db 514 AALLARIGDSTFYEDAAEFGLLPPEVRYFRFGPLYANKDF-----LRSLXSLGLD 569
Qy 562 QDEENAPPTDNNQTPANGTSVSYTFSPDSSPPAOSPEPPASABAPGEPDMLASVPP 621
Db 570 AGT-SATRKRDGTEVGSNSRLVDRKDLGSVSS-----GDGLV-VPLA 610
Qy 622 VTHHTLLIDMGSGVSPVDLMGKALAKLSITGKIGKVFVNIHAQYINDISHGVE-ED 680
Db 611 FGHHTVVIDCAPLFLFDVAGMATLKLRLKNRYRALDITLLACCCSPVDTLRKGGFLGED 670
Qy 681 GSLECKHVPFSIHDAVLEAQNARADV 706
Db 671 QGTAEELLFPVSHSAVETACARREL 696

RESULT 7
YLDM CAEEL
ID YLDM CAEEL STANDARD: PRT: 758 AA.
AC Q94225;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOHETICAL 85.0 KDA PROTEIN F41D9.5 IN CHROMOSOME X.
GN F41D9.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Fulton B.;

```


FT TRANSMEM 383 403 POTENTIAL.
 FT TRANSMEM 420 440 POTENTIAL.
 FT TRANSMEM 447 467 POTENTIAL.
 FT TRANSMEM 481 501 POTENTIAL.
 SO SEQUENCE 662 AA; 72728 MW; 7544DLE73FB76284 CRC64;

Query Match 13.4%; Score 518; DB 1; Length 662;
 Best Local Similarity 26.0%; Pred. No. 3e-27;
 Matches 183; Conservative 121; Mismatches 277; Indels 122; Gaps 22;

18 TLPDDEFEKKDRY----PVGEKLNARFRCSSAKIKAVFGL---LPVLSMLPKYIKDY 70
 Db TLFOELKHSNFEFFPKDQSGFR-----FKDQSSRK-FVLGLQYIRPILDMGRHNDLKR 101
 71 IIPDLGSGSGSIOVPGMAFALLANLPAVNGLYSSFFPLITFFLGGVHOMVPGFAV 130
 Db 102 -KGFDFAGLTITSLICIPODIATAKLANLDPWGLYSSFAVPLVAFMGTSRDIAIGFAV 160
 131 ISLVGNICLOLAPESKFOVFNATNESYVDTAAMEAEHLVASATLACTAIIONGLGFM 190
 Db 161 VSLILGTL---LSNE-----ISNTKSHDYL-----RLAFATF--FAGVTOMLLGVC 202
 191 QGFVAIYIYSESFIRGFMFAAGLIQILISVKYIFGLTIPSYGPGSIVTFPIDCKNLPH 230
 Db 203 RLGFLLDFLSHAIVGFMAAGAITIGLOOLKGLG--IKDFTKNSDIVSMHVSVMVHH 260
 251 -TNIASLIPALSGAFVLVYKELNARYMKIRPIPTEMIVVAVATLISGCGMKPKYHM 309
 Db 261 GNMWETILGLSLFLILTKYIAKNNKLEFWSAISPMICVISTFEVYITRADKR-GV 319
 310 QIVGEIORGF-PTPVSPVVSOKMDIGTAFSLAIVSYNL---AMGRTLANKHGYDND 364
 Db 320 TIVKHIKSGVNSANEIFFHGR-KYLGAGVRGVAVAGVALTEPAMAIGTFPAMKDYSD 378
 365 SNOEMALICGSNFGSFRIHVICALSVTLAVDAGGKSQVSLCVSLVMTMLVGLI 424
 Db 379 GKNEMVAMGTMTNIVGSLTSCYVTGFSRSAAVVMAGCKTAIVNIAVAILTLLVITP 438
 425 YLYPLKSVLGLI---AVNLKNSLKOLDPYLLMRKSLDCCIWVVSFLSFFLSPLYG 461
 Db 439 LKRYTNNAYLASSIILAAVNVLVN-----IEAMVLLMKIDKDFPACMGAFGCVIFKSYEIG 494
 482 VAVGAFSVLVVVFQTFRNGYALAOVMDTIVNPKTYNRAODLOGIKLITVCSPLYFA 541
 Db 495 LILAVIASPAKILLYATRPRTAVLGKLPSTSVYRNIOQYPKAQIDPMLIIRVDSAIYFS 554
 542 NSRIFQOKYIAKVSIQELQODFENAPPTDPNNNOTPANGTSVYITFSDSSPAQSEP 601
 Db 555 NSMYIKERILMLID-----EGAQRTSESL 579
 602 PASAEPARPSMDLAVPPEVTFHTLLDMGVSFVDMGKALKALKSTYYGKIGVFL 661
 Db 580 P-----ETIOHLIVEMSPYTDIDTDSIHAFELIKYLQKREVOLML 619
 662 VN-----HAQVYNDISHGVEFEDGSLCKHVPESIDAV 696
 Db 620 ANGPVYIEKLHASNLAEL-----IGED-----KIFLVADAV 652

RESULT 9
 SUTL_STVHA STANDARD; PRT; 667 AA.
 AC P53391;
 DT 01-OCT-1996 (rel. 34, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 01-OCT-1996 (rel. 34, Last annotation update)
 DE HIGH AFFINITY SULPHATE TRANSPORTER 1.
 GN STL.
 OS Stylosanthes hamata.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Speimacophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;

OC Stylosanthes.
 NCBI_TaxID=37660;
 ON (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. VERRANO. TISSUE=Root;
 RX MEDLINE=96016171; PubMed=7568135;
 RA Smith F.W., Ealing P.M., Hawkesford M.J., Clarkson D.T.;
 RT "Plant members of a family of sulfate transporters reveal functional
 subtypes";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:9373-9377(1995).
 CC - FUNCTION: HIGH-AFFINITY H+/SULFATE COTRANSPORTER THAT MEDIATES THE
 UPTAKE OF SULFATE BY PLANT ROOTS FROM LOW CONCENTRATIONS OF
 SULFATE IN THE SOIL SOLUTION.
 CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC - SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; X82255; CAA57710.1; -
 CC InterPro: IPR002645; STAS.
 CC InterPro: IPR001902; Sulfate_transp.
 CC Pfam: PF00916; Sulfate_transp. 1.
 CC PROSITE: PS01130; SULFATE_TRANS. 1.
 DR Transmembrane; Transport; Multigene family.
 KW TRANSMEM 106 126 POTENTIAL.
 FT TRANSMEM 131 151 POTENTIAL.
 FT TRANSMEM 156 176 POTENTIAL.
 FT TRANSMEM 185 205 POTENTIAL.
 FT TRANSMEM 208 228 POTENTIAL.
 FT TRANSMEM 269 289 POTENTIAL.
 FT TRANSMEM 296 316 POTENTIAL.
 FT TRANSMEM 350 370 POTENTIAL.
 FT TRANSMEM 425 445 POTENTIAL.
 FT TRANSMEM 452 472 POTENTIAL.
 FT TRANSMEM 486 506 POTENTIAL.
 FT SEQUENCE 667 AA; 73172 MW; 1A7AD47FDEB4DA7E CRC64;

Query Match 13.3%; Score 516; DB 1; Length 667;
 Best Local Similarity 25.5%; Pred. No. 4.1e-27;
 Matches 179; Conservative 128; Mismatches 278; Indels 116; Gaps 20;

18 TLPDDEFEKKDRY---PVGE-KLNARFRCSSAKIKAVFGLLPVLSMLPKYIKINDYII 72
 Db TLFOELKHSNFEFFPKDQSGFR-----KLELGLQYIRPILDMGRHNDLKR-R 105
 73 PDLGSGSGSIOVPGMAFALLANLPAVNGLYSSFFPLITFFLGGVHOMVPGFAVIS 132
 Db 106 GDFPAGLTITSLICIPDIAATAKLANLDPWGLYSSFAVPLVAFMGTSRDIAIGFAVVS 165
 131 ISLVGNICLOLAPESKFOVFNATNESYVDTAAMEAEHLVASATLACTAIIONGLGFMQF 192
 Db 166 LILGTL---LSNE-----ISNTKSHDYL-----RLAFATF--FAGVTOMLLGVCRL 207
 193 GFAVAYIYSESFIRGFMFAAGLIQILISVKYIFGLTIPSYGPGSIVTFPIDCKNLPH-T 251
 Db 208 GLFDLDFLSHAIVGFMAAGAITIGLOOLKGLGISTNNFRTKFDIISVMSVMTVHHNGW 267
 252 NIASLIFALISGAFVLVYKELNARYMKIRPIPTEMIVVAVATLISGCGMKPKYIMO 311
 Db 268 NMWETILGLSFLFLITTKYIAKNNKLEFWSAISPMICVISTFEVYITRADKR-GVSI 326
 312 VGEIORGF-PTPVSPVVSOKMDIGTAFSLAIVSYNL---AMGRTLANKHGYDVDSN 366
 Db 327 VKHIKSGVNSANEIFFHGR-KYLGAGVRGVAVAGVALTEPAMAIGTFPAMKDYALDGN 385

QY 367 QEMALCCSNEFSGFKIHVICALSVTLAVDAGGKSQVASCVSLVMTMLVLGILY 426
 Db 386 KEMVAMGTMTNIVGSLSSCYVTGTSFSSAVVYAGCKTAVSNVIMSVILLTLVTLPLF 445
 QY 427 YPLKSVGLAL---ANLKSLSKQLDPPYLYMKRSKIDCCIMVVSFLSFLSLPYVA 483
 Db 446 KYTPNAVLASTIIIAVNLVN---IEAMVLLMKIDKFDVACGAFEGVIFKSVIEGLL 501
 QY 484 VGVAFLVAVVFOQFRNGYALQVMDTIVNPKTYNRAODIGIKITTCSPLYEANS 543
 Db 502 IAVNISFAKILLOVTRPRTAVLGKLPQSVYRNIOQYPKAQIGMLIRDSAIYENS 561
 QY 544 EIFQOKYIAKTVSLQELQODEFNAPTDPNNOTPANGTSVYITTFSDSSPAOSEPPA 603
 Db 562 NYIERILRLMID-----EGAQRTESELP- 585
 QY 604 SAAPGEPDMLASVPFVFHTLLDMGVSFVDMGKALAKLSSTYGGIKVFLVN 663
 Db 586 -----ELQHLITTEKSPDPDIDTSGIHAFEEUYKTLQREVOLILAN 626
 QY 664 -----IHAQVYNDISHGCVFEDGSLCKHVPFSIHDAV 696
 Db 627 PGVPIEKLNHASKLTCL-----KIFLYADAV 657

RESULT 10
 SUT3_STYHA STANDARD; PRT; 644 AA.
 AC P53393;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE LOW AFFINITY SULPHATE TRANSPORTER 3.
 GN ST3.
 OS Stylosanthes hamata.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;
 OC Stylosanthes.
 ON NCBI_TaxID=37660;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CY. VERANO; TISSUE=Root;
 RX MIMDLINE=96016171; PubMed=7568135;
 RA Smith F.W., Ealing P.M., Hawesford M.J., Clarkson D.T.;
 RT "Plant members of a family of sulfate transporters reveal functional
 subtypes";
 RT Proc. Natl. Acad. Sci. U.S.A. 92:9373-9377(1995).
 CC -1- FUNCTION: LOW-AFFINITY H+/SULFATE COTRANSPORTER WHICH MAY BE
 INVOLVED IN THE INTERNAL TRANSPORT OF SULFATE BETWEEN CELLULAR OR
 SUBCELLULAR COMPARTMENTS WITHIN THE PLANT.
 CC -1- SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X82454; CAAS7831.1; -
 DR InterPro: IPR002645; STAS.
 DR InterPro: IPR001902; Sulfate_transp.
 DR Pfam: PF01740; STAS; 1. Sulfate_transp.
 DR Pfam: PF00916; Sulfate_transp; 1.
 DR PROSITE: PS0130; SULFATE_TRANS; 1.
 KW Transmembrane; Transport; Multigene family.
 FT TRANSMEM 50 70 POTENTIAL.
 FT TRANSMEM 76 96 POTENTIAL.
 FT TRANSMEM 99 119 POTENTIAL.
 FT TRANSMEM 124 144 POTENTIAL.

FT TRANSMEM 156 176 POTENTIAL.
 FT TRANSMEM 179 199 POTENTIAL.
 FT TRANSMEM 242 262 POTENTIAL.
 FT TRANSMEM 268 288 POTENTIAL.
 FT TRANSMEM 328 348 POTENTIAL.
 FT TRANSMEM 394 414 POTENTIAL.
 FT TRANSMEM 418 438 POTENTIAL.
 FT TRANSMEM 455 475 POTENTIAL.
 FT TRANSMEM 518 538 POTENTIAL.
 SO SEQUENCE 644 AA; 69785 MW; DA2A0D4FDB5ED280 CRC64;

Query Match 12.88; Score 496; DB 1; Length 644;
 Best Local Similarity 26.48; Pred. No. 8.5e-26;
 Matches 180; Conservative 127; Mismatches 260; Indels 116; Gaps 26;

QY 41 FRCSAK--IKAYVE--GLPLVSLMPKY---KIKDYIIPPLGLSGSIOVQGNMA 93
 Db 39 FTSSSKKETRAVSFLASLPILSMIRYSATKRD---DLISGLTLASLIFQSIGYA 94
 QY 94 LLALPANGLSSEFFPLTYEFLGVHQMVGTFANVISILVGNICQLAPESKFOYENN 153
 Db 95 NLAKLDPQGLTYSVIPVYIALMGSSREIAGPAAVSMLLSS---LVPK----- 142
 QY 154 ATNESYDTAMEARELHVSATLACLPATIDMGICFMQFGVAIYLSSEFRTAGL 213
 Db 143 -----VIDPDAPHNDYRNLVFTVTLFAGIFQTAGVLEGLVDFLSHAALVGMAGAI 197
 QY 214 QILSVLKTYRIGLTPSYTGPISVFPFIDCKMLPHNINAS-----LIPALISGAFV 267
 Db 198 VIGLQKLGGLT--HETTDVAVNLKSVYTSI-HQOITSENWSLNF-VIGCSFLI 253
 QY 268 LKLELMARYM--HKIREPIP--TEMIVVVATAI---SGCKMKPKYHMOIVGEIQGF 319
 Db 254 FL--LAARFIRRNKKKFFMLPALPILSLVILSTLIVFLSKG---DKHGVNIIKHVGG 307
 QY 320 -PTVSPVYSGMKMITGASLAIVSYINL---AMGRTLANKHGVDVDSNCKMALGC 374
 Db 308 NPSSVHKIQLNGPH-VGQAAKIGLISALITLTAIAGRSFANKGYHLDNKEMLAMGC 366
 QY 375 SNFGSFFKIHVICALSVTLAVDAGGKSQVASCVSLVMTMLVGLYLYPLKPSVL 434
 Db 367 MNINGSLSCTVSGTSRSRTAVNFSAGCKTAVSNVIMAVYLLCLELFTLLYTPPAITL 426
 QY 435 GALLAVNLKLSKQLDPPYLYMKRSKIDCCIMVVSFLSFLSLPYGAVGAVSVLVV 494
 Db 427 ASILSLALP-GLIDIGEAHIMKVKDEPLACLAFGLVFLVSIEIGLIALSIPAKIL 485
 QY 495 FQTFRNGYALQVMDTIVNPKTYNRAODIGIKITTCOS-PLYFRANSEIFQOKYIAK 553
 Db 486 LQAIRPGVEVLGRIPTEAYCDVQAQYPMAYVTPGILVIRISSGSLCFAMAGFVREIRL-K 544
 QY 554 TVSLQELQODEFNAPTDPNNOTPANGTSVYITTFSDSSPAOSEPPAASEGERSD 613
 Db 545 WVEDEE-ODNIEEA-----AKG----- 560
 QY 614 MLASVPFVFHTLLDMGVSFVDMGKALAKLSSTYGGIKVFLVNIHAQVYNDIS 673
 Db 561 -----RQALIIIDMTDLTNDTSGIILAEHLKHLKLSRGVLAAMVNRMEVHKIL 611
 QY 674 HGVFEDGSLCKHVPFSIHDAV 696
 Db 612 VAN-FVD-KIGKERVFLTVAEAV 632

RESULT 11
 SUT3_STYHA STANDARD; PRT; 877 AA.
 AC 074377;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROBABLE SULFATE PERMEASE C3H7.02.

```

Db 319 LKRTKTRQRFVFLNLNLRSAVILIVGALISYGVCKHKHREMPISLIGTVPSGRDMGVP 378
QY 326 VVSQ--WKMDIGTAFSLAIVSYVINLNGRPLANKHGVDVDSNODMTALGCSNFGSEFEK 383
Db 379 VISRKLACADLASELPVSVIYVLLLEHISIAKSFGRVNDKYVPIDDELLAMGATNLIGVFFH 438
QY 384 IHVICCALSVLAVDAGAGKSGQVASICSVLVMTIMLVGLIYLPPLKSVGALLIANVLK 443
Db 439 AVPATGFSFSRAINAKSGSVTRPPLGGIFTAGVVALVLCYLTGAFFIYIPNAVLSAVIHSVF 498
QY 444 NSLKQLTDPYVLMKRSKIDCCIVWVSFLTSEFFLSLPYGVAVGVAFSVLVVFOFOFRNG- 502
Db 499 DLIPMRQRTLTFWKRQPLLEALIFCAVVSVFSSITENIYTAVCLSAALLFRIAKPSGS 558
QY 503 ---YALAQVND-----TDIVV--NPKTYNRADIOGKIITYCSPLYFANSEIFR 547
Db 559 FLGLIKIANKFDDDDENSIDVVRDIYPLNOKGMN-----PNLTVRDPP--AGVLIFR 608
QY 548 QKVIKATYVLSLOEODFENAPPTDPNNNOTRANGS-VSYIT-----FSPDSSSPQOS 599
Db 609 -----LQESF-----TYPNAGHVNSMLTSKAKTYTRRCGNANIYRKASDRPWND 651
QY 600 EPPAAEAEPGEPSDMLASVPPEVFRHTLILDMSGVSPYDLMLGIRKLA--KLSTPYGKIG 656
Db 652 PAPRKKKNAPEYED-----TRPLL--RAIILDFSAVNHIDTGVQALVYTRKELEYADE 705
QY 657 VKVFLVNIHAQVYND-----ISHG-GVEEDGSLCKKHVPFSIHDAVLFAQAMARDV--- 706
Db 706 VEFHFTDIN---NDMIKRTLVAAGFGKARAT--KYTSRISIE---VGSAAPLRDIETP 755
QY 707 -TPGHNFGCAGDAELSLYDSEDIORS 732
Db 756 MAPGNSRIWMPSSVVRPEFDEEALES 782

RESULT 12
SULX_YEAST
ID SULX_YEAST STANDARD; PRT; 754 AA.
AC P53394;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PUTATIVE SULFATE TRANSPORTER YPR003C.
GN YPR003C OR YP9723.03C OR LP23C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxId=4932;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Badcock K., Bowman S., Churcher C.M., Pearson D., Rajandream M.A.,
RA Walsh S.V., Barrell B.G.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: POSSIBLE SULFATE TRANSPORTER.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z71255; CAAG5043.1; -
CC EMBL; U31900; AAA97582.1; -
CC EMBL; Z48951; CAAB8781.1; -
CC SGD; S0006207; YPR003C.
CC InterPro; IPR002645; STAS.
CC InterPro; IPR001902; Sulfate_transp.
CC Pfam; PF01740; STAS; 1.

```

DR Pfam: PF00916; Sulfate_transp: 1.
 DR PROSITE: PS01130; SULFATE_TRANS: FALSE_NEG.
 KW Hypothetical protein; Transmembrane; Transport.
 FT TRANSMEM 119 139 POTENTIAL.
 FT TRANSMEM 147 167 POTENTIAL.
 FT TRANSMEM 173 193 POTENTIAL.
 FT TRANSMEM 200 220 POTENTIAL.
 FT TRANSMEM 233 253 POTENTIAL.
 FT TRANSMEM 283 303 POTENTIAL.
 FT TRANSMEM 318 338 POTENTIAL.
 FT TRANSMEM 371 391 POTENTIAL.
 FT TRANSMEM 411 431 POTENTIAL.
 FT TRANSMEM 451 471 POTENTIAL.
 FT TRANSMEM 475 495 POTENTIAL.
 FT TRANSMEM 518 538 POTENTIAL.
 SQ SEQUENCE 754 AA; 83383 MW; 4558DAB879803EEG CRC64;

Query Match 11.38; Score 437; DB 1; Length 754;
 Best Local Similarity 22.5%; Pred. No. 9e-22;
 Matches 162; Conservative 138; Mismatches 265; Indels 154; Gaps 23;

QY 56 LPLVSLWPKKIKOYIIPDLGLSGSGSIQVPGMAFAL-LANLPAVNGLYSSFFLLTY 114
 DB 105 LPEFSWLPETFE-NKLMGDVIAGISVASFQIPALSTYTAHPPDLGSLAISPEVY 163
 QY 115 FFLGGVQWVPGTEFAVSIIVGNICQLAPESKFOVFNATNESYVDFAEAERLHVA 174
 DB 164 GILGSVQWVIGPESALSIVGQ-----AVES---ITLKHENSLID-----IST 205
 QY 175 TLACLAIIOMGLGEMOGEVAVALYLSFSGFRGEMTAAGLOLISLVL-----KYIFGLT 227
 DB 206 VIFVFSSTILLFSGISFRGELGNLAKLRFISVGLVMIINLSLEKLDKFLVSLP 265
 QY 228 IPSTGSGIVFPIIDICKNLPHNTIASLFPALISGAFVIVKRLNRYM--HKIRFPI 285
 DB 266 -QHHPPEKILFLIDAPAQYH--IPTALFSGCCLIVLEFLTRLLKRLMKMYHSAIFP 322
 QY 286 TEMIVVAVATAISGCGMKPKYHMQVGEIQR-GFPTPVSVVSOQWMDIGTARS---L 340
 DB 323 DILLVIVITLISMKFKLHKRYGSIIGDPSMDNDELKNLTPRRKLLPDLFSASLIY 382
 QY 341 AIVSYVNTLAKGRLANKHGYDSDNOEMIALGCSNFFGSFF-----KIHVIC 388
 DB 383 AMGFESTTASKSLGTTYNLTVSSNRELVALGFMNTIVISLFGALPAFGYGRSKINAL- 441
 QY 389 CALSVTLAVDAGGKSOVASLGVMTLMVIGIYLPKSVLALLAVNLKNSLKQ 448
 DB 442 -----SGAQSVMGVEKVIITLTMNLLOFVHIIPNCVLSVITTTIGISLDEE 490
 QY 449 LTPDYVLMRSKSLDC-----CIWVVSFLSFFLSPFYGAAGVAFSVLVVFPOTOPRNG 502
 DB 491 VPGDI-----KFHLRCGGFSELPFAVTEGTTIFYSIAGICIGCVYSIINLIKMSASRI 546
 QY 503 YALAQVMDTIYVNPRTY-----NRAODIGIKITTYSPLYFANSEIFRQKVI 551
 DB 547 QILARVAGTGNFTNLDYMMNMRNSLDVGEIEGCMIVRLEPELTFNSELDELQRL- 605
 QY 552 AKTVSLQELQDFENAPPTDPNNQTPANGTSVYIFPSDSSSPAQSEPPASAEAGEP 611
 DB 606 -----DRIERYGSSKIH-----PGRK 621
 QY 612 S-----DLASVPPFVTFHTLILDMGSGVSVFDLMGKALAKLSYTGKIGKVFVNTHAQ 667
 DB 622 SLRSKDSIKRYV-----IFDLGWTSIDSSAAQVLEELIITSYKRRNVFIYLVNS-- 670
 QY 668 VYVDISHGVFEEDGSLCKKAVFPISINDAVLFAQANARDVTPGHNFGCAPGDDELSTLYDS 726
 DB 671 -INDKVRRLFKAG-----VAASVERAQANNMNMNNTSNTFSDA-GETVSPYDS 717

ID SUL2_YEAST STANDARD; PRT; 893 AA.
 AC Q12325;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE SULFATE PERMEASE 2 (HIGH-AFFINITY SULFATE TRANSPORTER 2).
 GN SUL2 OR SEL2 OR YLR092W OR U9449.1.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C / AB972;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
 RA Favell A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,
 RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
 RA Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,
 RA Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,
 RA Tatch A., Trevasakis E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,
 RA Wilson R., Waterston R.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Benes V., Rechmann S., Nentwich U., Schwager C., Ansoerge W., Voss H.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=97207837; PubMed=9055073;
 RA Cherest H., Davidian J.C., Thomas D., Benes V., Ansoerge W.,
 RA Surdin-Kerjan Y.;
 RT "Molecular characterization of two high affinity sulfate transporters
 RT in Saccharomyces cerevisiae.";
 RL Genetics 145:627-635(1997).
 CC - FUNCTION: HIGH AFFINITY UPTAKE OF SULFATE INTO THE CELL.
 CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC - SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.
 CC - This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U53880; AAB67596.1; -
 DR EMBL: Z73264; CAA97653.1; -
 DR EMBL: Z73265; CAA97653.1; -
 DR EMBL: U53876; AAB67550.1; -
 DR SGD: S0004082; SUL2.
 DR InterPro: IPR002645; STAS.
 DR InterPro: IPR001902; Sulfate_transp.
 DR Pfam: PF01740; STAS; 1.
 DR Pfam: PF00916; Sulfate_transp: 1.
 DR PROSITE: PS01130; SULFATE_TRANS: 1.
 KW Transport; Transmembrane; Glycoprotein.
 FT TRANSMEM 132 152 POTENTIAL.
 FT TRANSMEM 164 184 POTENTIAL.
 FT TRANSMEM 189 209 POTENTIAL.
 FT TRANSMEM 222 242 POTENTIAL.
 FT TRANSMEM 245 265 POTENTIAL.
 FT TRANSMEM 306 326 POTENTIAL.
 FT TRANSMEM 351 371 POTENTIAL.
 FT TRANSMEM 400 420 POTENTIAL.
 FT TRANSMEM 444 464 POTENTIAL.
 FT TRANSMEM 484 504 POTENTIAL.
 FT TRANSMEM 539 559 POTENTIAL.
 FT TRANSMEM 560 580 POTENTIAL.
 SQ SEQUENCE 893 AA; 99650 MW; 67826955ACTC0BF5 CRC64;

Query Match 10.5%; Score 405; DB 1; Length 893;

Best Local Similarity 23.3%; Pred. No. 1,5e-19;
Matches 157; Conservative 118; Mismatches 240; Indels 160; Gaps 28;

```

OY 17 LTLFDEFEK-KDRTPYVGEKLRNNAFR-----CSSAKIAVVGGLPVSIMLKYKI 67
DB 81 VSLYNSVSQFEETVYSLEKYDHSIRSLHYKACSYLK-----SVPIIMLPHYNF 134
OY 68 KDYIIPDLGLSGSIOVPGMAFALLANLPAVNGLYSSFPPLTYFFLGWHQVPGT 127
DB 135 -SMFADLAGITICCVLPQSMASYAQAVALTPAQYGLSYSSFGAYSYSPFANSKDVCI 193
OY 128 FAVISILVNICLOLARPEKFOVFNNAATNESYVDPRAMAERLHVSATLACTAIIOK 187
DB 194 VAWMSLQAKAVIADVT--AKYPDGSALTGPVIAT-----TALLCGIISAAY 239
OY 188 GPMQGFVAIVYSESPFIRGFMFAAGLOILSVLYKIFGLTIPSTPGSIVFT----- 240
DB 240 GFLRLGFLVELISLNAVGFMTGSANIL-----WG-QVALMKYNSLVNTRATYKV 291
OY 241 FIDICKNLPHNINIASLIFALISGAFVLV-----KELNARY-----MHKI----- 280
DB 292 VLETKHLDPDKL-DAVEGLIP-LFLLYWKWCGTYGPRLNDRYNSKNRPLKIKMTY 349
OY 281 -RFPITFIVVAVT-----AISGCKMKKXHMQIVGEIQGF-PTPSPVYSQKMDI 334
DB 350 FPAQASRNGIITIVFCIGMALTREG-KSKSERPISILGVSPLKEVGFVHPGLMSKL 408
OY 335 GAFSLAIVSYVI-NLAMQRTLANKHGYVDNSNOEMIALGCSNFSGFKIHVICALSV 393
DB 409 GNLPAISIVLLEHIAISKSGRIIDYKVPDQELIAGVSNLGTFFINAY----- 460
OY 394 TLAVDAGCKSOVASLC-----VSLVMTMLVGLIYLPKPSVGCALIAVNL 442
DB 461 --PATGFSFSAKAKCNVTRPLSLGSLGSCVLLALCYCTGAFFY-IPRATLSAVIHA 517
OY 443 KASLQGLTPYTLMRKSKIDCCIMWVVSFLSPFLSPGVANAVASVLYVQGFRRNG 502
DB 518 SULLASVQTTMFMKKNPLDFTICFIVTALITVFASTEDDIYRAMCSCAMLLIKVAF 577
OY 503 VALAQV---MDTDIYVNPXTYNRAODIQGKITTCSPLYFANSELFROKVIKTVSLO 559
DB 578 KLGREVAEVDATVRRPS-----DVSYVS----- 604
OY 560 LQODEFNAEPPTPNNN--OTPANG-----TSVSYTSPDSSSPAQSEPPASAEAP 611
DB 605 -----ENNNGISTLEDGEDKESSTKYVT-----NSSKIIETNQTGKFDSP 647
OY 612 SDMLASVPEFVTEHT 626
DB 648 SSSISO--PRIKYHT 660

```

RESULT 14
SUL1_YEAST
ID SUL1_YEAST STANDARD; PRT; 859 AA.
AC P38359;
DT 01-OCT-1994 (Rel. 30, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE SULFATE PERMEASE 1 (HIGH-AFFINITY SULFATE TRANSPORTER 1).
GN SUL1 OR SFP OR YBR294W OR YBR2110.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INVSCL;
RA MEDLINE=95342164; PubMed=7616962;
RA Smith F.W., Hawksworth M.J., Prosser I.M., Clarkson D.T.;
RT "Isolation of a cDNA from Saccharomyces cerevisiae that encodes a
Mol. Genet. 247:709-715(1995)."

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RA Fritz C., Hollenberg C.P., Kirchath L., Rad M.R.;
 RL Submitted (Mar-1998) to the EMBL/Genbank/DDJ databases.
 RN [3]
 RP SEQUENCE OF 1-585 FROM N.A.
 RC STRAIN=JP1721-1B, AND DB747;
 RA Jin Y.H., Jang Y.K., Kim M.J., Lee J.S., Huh S.O., Hong S.H.,
 RA Park S.D.;
 RL Submitted (JUL-1994) to the EMBL/Genbank/DDJ databases.
 RN [4]
 RP CHARACTERIZATION.
 RA MEDLINE=97207837; PubMed=9055073;
 RA Cherest H., Davidian J.C., Thomas D., Ansoorge W.,
 RA Sardin-Kerjan Y.;
 RT "Molecular characterization of two high affinity sulfate transporters
 in Saccharomyces cerevisiae.";
 RL Genetics 145:627-633(1997).
 CC -1- FUNCTION: HIGH AFFINITY UPTAKE OF SULFATE INTO THE CELL.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.
 CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
 FRAMESHIFT IN POSITION 488.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: Z36163; CAA85259.1; -
 DR EMBL: X82013; CAA57540.1; -
 DR EMBL: Z35134; CAA84506.1; ALT_FRAME.
 DR PIR: S46176; S46176.
 DR SGD: S0000498; SUL1.
 DR InterPro: IPR002645; STAS.
 DR InterPro: IPR001902; Sulfate_transp.
 DR Pfam: PF01740; STAS: 1.
 DR Pfam: PF00916; Sulfate_transp.
 DR PROSITE: PS01130; Sulfate_transp. 1.
 KW Transport; Transmembrane; Glycoprotein.
 FT TRANSMEM 94 114
 FT TRANSMEM 116 136
 FT TRANSMEM 148 168
 FT TRANSMEM 173 193
 FT TRANSMEM 206 226
 FT TRANSMEM 234 254
 FT TRANSMEM 292 312
 FT TRANSMEM 332 352
 FT TRANSMEM 395 415
 FT TRANSMEM 428 448
 FT TRANSMEM 468 488
 FT TRANSMEM 525 545
 FT TRANSMEM 545 565
 FT TRANSMEM 585 605
 FT TRANSMEM 625 645
 FT TRANSMEM 665 685
 FT TRANSMEM 705 725
 FT TRANSMEM 745 765
 FT TRANSMEM 785 805
 FT TRANSMEM 825 845
 FT TRANSMEM 865 885
 FT TRANSMEM 905 925
 FT TRANSMEM 945 965
 FT TRANSMEM 985 1005
 FT TRANSMEM 1025 1045
 FT TRANSMEM 1065 1085
 FT TRANSMEM 1105 1125
 FT TRANSMEM 1145 1165
 FT TRANSMEM 1185 1205
 FT TRANSMEM 1225 1245
 FT TRANSMEM 1265 1285
 FT TRANSMEM 1305 1325
 FT TRANSMEM 1345 1365
 FT TRANSMEM 1385 1405
 FT TRANSMEM 1425 1445
 FT TRANSMEM 1465 1485
 FT TRANSMEM 1505 1525
 FT TRANSMEM 1545 1565
 FT TRANSMEM 1585 1605
 FT TRANSMEM 1625 1645
 FT TRANSMEM 1665 1685
 FT TRANSMEM 1705 1725
 FT TRANSMEM 1745 1765
 FT TRANSMEM 1785 1805
 FT TRANSMEM 1825 1845
 FT TRANSMEM 1865 1885
 FT TRANSMEM 1905 1925
 FT TRANSMEM 1945 1965
 FT TRANSMEM 1985 2005
 FT TRANSMEM 2025 2045
 FT TRANSMEM 2065 2085
 FT TRANSMEM 2105 2125
 FT TRANSMEM 2145 2165
 FT TRANSMEM 2185 2205
 FT TRANSMEM 2225 2245
 FT TRANSMEM 2265 2285
 FT TRANSMEM 2305 2325
 FT TRANSMEM 2345 2365
 FT TRANSMEM 2385 2405
 FT TRANSMEM 2425 2445
 FT TRANSMEM 2465 2485
 FT TRANSMEM 2505 2525
 FT TRANSMEM 2545 2565
 FT TRANSMEM 2585 2605
 FT TRANSMEM 2625 2645
 FT TRANSMEM 2665 2685
 FT TRANSMEM 2705 2725
 FT TRANSMEM 2745 2765
 FT TRANSMEM 2785 2805
 FT TRANSMEM 2825 2845
 FT TRANSMEM 2865 2885
 FT TRANSMEM 2905 2925
 FT TRANSMEM 2945 2965
 FT TRANSMEM 2985 3005
 FT TRANSMEM 3025 3045
 FT TRANSMEM 3065 3085
 FT TRANSMEM 3105 3125
 FT TRANSMEM 3145 3165
 FT TRANSMEM 3185 3205
 FT TRANSMEM 3225 3245
 FT TRANSMEM 3265 3285
 FT TRANSMEM 3305 3325
 FT TRANSMEM 3345 3365
 FT TRANSMEM 3385 3405
 FT TRANSMEM 3425 3445
 FT TRANSMEM 3465 3485
 FT TRANSMEM 3505 3525
 FT TRANSMEM 3545 3565
 FT TRANSMEM 3585 3605
 FT TRANSMEM 3625 3645
 FT TRANSMEM 3665 3685
 FT TRANSMEM 3705 3725
 FT TRANSMEM 3745 3765
 FT TRANSMEM 3785 3805
 FT TRANSMEM 3825 3845
 FT TRANSMEM 3865 3885
 FT TRANSMEM 3905 3925
 FT TRANSMEM 3945 3965
 FT TRANSMEM 3985 4005
 FT TRANSMEM 4025 4045
 FT TRANSMEM 4065 4085
 FT TRANSMEM 4105 4125
 FT TRANSMEM 4145 4165
 FT TRANSMEM 4185 4205
 FT TRANSMEM 4225 4245
 FT TRANSMEM 4265 4285
 FT TRANSMEM 4305 4325
 FT TRANSMEM 4345 4365
 FT TRANSMEM 4385 4405
 FT TRANSMEM 4425 4445
 FT TRANSMEM 4465 4485
 FT TRANSMEM 4505 4525
 FT TRANSMEM 4545 4565
 FT TRANSMEM 4585 4605
 FT TRANSMEM 4625 4645
 FT TRANSMEM 4665 4685
 FT TRANSMEM 4705 4725
 FT TRANSMEM 4745 4765
 FT TRANSMEM 4785 4805
 FT TRANSMEM 4825 4845
 FT TRANSMEM 4865 4885
 FT TRANSMEM 4905 4925
 FT TRANSMEM 4945 4965
 FT TRANSMEM 4985 5005
 FT TRANSMEM 5025 5045
 FT TRANSMEM 5065 5085
 FT TRANSMEM 5105 5125
 FT TRANSMEM 5145 5165
 FT TRANSMEM 5185 5205
 FT TRANSMEM 5225 5245
 FT TRANSMEM 5265 5285
 FT TRANSMEM 5305 5325
 FT TRANSMEM 5345 5365
 FT TRANSMEM 5385 5405
 FT TRANSMEM 5425 5445
 FT TRANSMEM 5465 5485
 FT TRANSMEM 5505 5525
 FT TRANSMEM 5545 5565
 FT TRANSMEM 5585 5605
 FT TRANSMEM 5625 5645
 FT TRANSMEM 5665 5685
 FT TRANSMEM 5705 5725
 FT TRANSMEM 5745 5765
 FT TRANSMEM 5785 5805
 FT TRANSMEM 5825 5845
 FT TRANSMEM 5865 5885
 FT TRANSMEM 5905 5925
 FT TRANSMEM 5945 5965
 FT TRANSMEM 5985 6005
 FT TRANSMEM 6025 6045
 FT TRANSMEM 6065 6085
 FT TRANSMEM 6105 6125
 FT TRANSMEM 6145 6165
 FT TRANSMEM 6185 6205
 FT TRANSMEM 6225 6245
 FT TRANSMEM 6265 6285
 FT TRANSMEM 6305 6325
 FT TRANSMEM 6345 6365
 FT TRANSMEM 6385 6405
 FT TRANSMEM 6425 6445
 FT TRANSMEM 6465 6485
 FT TRANSMEM 6505 6525
 FT TRANSMEM 6545 6565
 FT TRANSMEM 6585 6605
 FT TRANSMEM 6625 6645
 FT TRANSMEM 6665 6685
 FT TRANSMEM 6705 6725
 FT TRANSMEM 6745 6765
 FT TRANSMEM 6785 6805
 FT TRANSMEM 6825 6845
 FT TRANSMEM 6865 6885
 FT TRANSMEM 6905 6925
 FT TRANSMEM 6945 6965
 FT TRANSMEM 6985 7005
 FT TRANSMEM 7025 7045
 FT TRANSMEM 7065 7085
 FT TRANSMEM 7105 7125
 FT TRANSMEM 7145 7165
 FT TRANSMEM 7185 7205
 FT TRANSMEM 7225 7245
 FT TRANSMEM 7265 7285
 FT TRANSMEM 7305 7325
 FT TRANSMEM 7345 7365
 FT TRANSMEM 7385 7405
 FT TRANSMEM 7425 7445
 FT TRANSMEM 7465 7485
 FT TRANSMEM 7505 7525
 FT TRANSMEM 7545 7565
 FT TRANSMEM 7585 7605
 FT TRANSMEM 7625 7645
 FT TRANSMEM 7665 7685
 FT TRANSMEM 7705 7725
 FT TRANSMEM 7745 7765
 FT TRANSMEM 7785 7805
 FT TRANSMEM 7825 7845
 FT TRANSMEM 7865 7885
 FT TRANSMEM 7905 7925
 FT TRANSMEM 7945 7965
 FT TRANSMEM 7985 8005
 FT TRANSMEM 8025 8045
 FT TRANSMEM 8065 8085
 FT TRANSMEM 8105 8125
 FT TRANSMEM 8145 8165
 FT TRANSMEM 8185 8205
 FT TRANSMEM 8225 8245
 FT TRANSMEM 8265 8285
 FT TRANSMEM 8305 8325
 FT TRANSMEM 8345 8365
 FT TRANSMEM 8385 8405
 FT TRANSMEM 8425 8445
 FT TRANSMEM 8465 8485
 FT TRANSMEM 8505 8525
 FT TRANSMEM 8545 8565
 FT TRANSMEM 8585 8605
 FT TRANSMEM 8625 8645
 FT TRANSMEM 8665 8685
 FT TRANSMEM 8705 8725
 FT TRANSMEM 8745 8765
 FT TRANSMEM 8785 8805
 FT TRANSMEM 8825 8845
 FT TRANSMEM 8865 8885
 FT TRANSMEM 8905 8925
 FT TRANSMEM 8945 8965
 FT TRANSMEM 8985 9005
 FT TRANSMEM 9025 9045
 FT TRANSMEM 9065 9085
 FT TRANSMEM 9105 9125
 FT TRANSMEM 9145 9165
 FT TRANSMEM 9185 9205
 FT TRANSMEM 9225 9245
 FT TRANSMEM 9265 9285
 FT TRANSMEM 9305 9325
 FT TRANSMEM 9345 9365
 FT TRANSMEM 9385 9405
 FT TRANSMEM 9425 9445
 FT TRANSMEM 9465 9485
 FT TRANSMEM 9505 9525
 FT TRANSMEM 9545 9565
 FT TRANSMEM 9585 9605
 FT TRANSMEM 9625 9645
 FT TRANSMEM 9665 9685
 FT TRANSMEM 9705 9725
 FT TRANSMEM 9745 9765
 FT TRANSMEM 9785 9805
 FT TRANSMEM 9825 9845
 FT TRANSMEM 9865 9885
 FT TRANSMEM 9905 9925
 FT TRANSMEM 9945 9965
 FT TRANSMEM 9985 10005
 FT TRANSMEM 10025 10045
 FT TRANSMEM 10065 10085
 FT TRANSMEM 10105 10125
 FT TRANSMEM 10145 10165
 FT TRANSMEM 10185 10205
 FT TRANSMEM 10225 10245
 FT TRANSMEM 10265 10285
 FT TRANSMEM 10305 10325
 FT TRANSMEM 10345 10365
 FT TRANSMEM 10385 10405
 FT TRANSMEM 10425 10445
 FT TRANSMEM 10465 10485
 FT TRANSMEM 10505 10525
 FT TRANSMEM 10545 10565
 FT TRANSMEM 10585 10605
 FT TRANSMEM 10625 10645
 FT TRANSMEM 10665 10685
 FT TRANSMEM 10705 10725
 FT TRANSMEM 10745 10765
 FT TRANSMEM 10785 10805
 FT TRANSMEM 10825 10845
 FT TRANSMEM 10865 10885
 FT TRANSMEM 10905 10925
 FT TRANSMEM 10945 10965
 FT TRANSMEM 10985 11005
 FT TRANSMEM 11025 11045
 FT TRANSMEM 11065 11085
 FT TRANSMEM 11105 11125
 FT TRANSMEM 11145 11165
 FT TRANSMEM 11185 11205
 FT TRANSMEM 11225 11245
 FT TRANSMEM 11265 11285
 FT TRANSMEM 11305 11325
 FT TRANSMEM 11345 11365
 FT TRANSMEM 11385 11405
 FT TRANSMEM 11425 11445
 FT TRANSMEM 11465 11485
 FT TRANSMEM 11505 11525
 FT TRANSMEM 11545 11565
 FT TRANSMEM 11585 11605
 FT TRANSMEM 11625 11645
 FT TRANSMEM 11665 11685
 FT TRANSMEM 11705 11725
 FT TRANSMEM 11745 11765
 FT TRANSMEM 11785 11805
 FT TRANSMEM 11825 11845
 FT TRANSMEM 11865 11885
 FT TRANSMEM 11905 11925
 FT TRANSMEM 11945 11965
 FT TRANSMEM 11985 12005
 FT TRANSMEM 12025 12045
 FT TRANSMEM 12065 12085
 FT TRANSMEM 12105 12125
 FT TRANSMEM 12145 12165
 FT TRANSMEM 12185 12205
 FT TRANSMEM 12225 12245
 FT TRANSMEM 12265 12285
 FT TRANSMEM 12305 12325
 FT TRANSMEM 12345 12365
 FT TRANSMEM 12385 12405
 FT TRANSMEM 12425 12445
 FT TRANSMEM 12465 12485
 FT TRANSMEM 12505 12525
 FT TRANSMEM 12545 12565
 FT TRANSMEM 12585 12605
 FT TRANSMEM 12625 12645
 FT TRANSMEM 12665 12685
 FT TRANSMEM 12705 12725
 FT TRANSMEM 12745 12765
 FT TRANSMEM 12785 12805
 FT TRANSMEM 12825 12845
 FT TRANSMEM 12865 12885
 FT TRANSMEM 12905 12925
 FT TRANSMEM 12945 12965
 FT TRANSMEM 12985 13005
 FT TRANSMEM 13025 13045
 FT TRANSMEM 13065 13085
 FT TRANSMEM 13105 13125
 FT TRANSMEM 13145 13165
 FT TRANSMEM 13185 13205
 FT TRANSMEM 13225 13245
 FT TRANSMEM 13265 13285
 FT TRANSMEM 13305 13325
 FT TRANSMEM 13345 13365
 FT TRANSMEM 13385 13405
 FT TRANSMEM 13425 13445
 FT TRANSMEM 13465 13485
 FT TRANSMEM 13505 13525
 FT TRANSMEM 13545 13565
 FT TRANSMEM 13585 13605
 FT TRANSMEM 13625 13645
 FT TRANSMEM 13665 13685
 FT TRANSMEM 13705 13725
 FT TRANSMEM 13745 13765
 FT TRANSMEM 13785 13805
 FT TRANSMEM 13825 13845
 FT TRANSMEM 13865 13885
 FT TRANSMEM 13905 13925
 FT TRANSMEM 13945 13965
 FT TRANSMEM 13985 14005
 FT TRANSMEM 14025 14045
 FT TRANSMEM 14065 14085
 FT TRANSMEM 14105 14125
 FT TRANSMEM 14145 14165
 FT TRANSMEM 14185 14205
 FT TRANSMEM 14225 14245
 FT TRANSMEM 14265 14285
 FT TRANSMEM 14305 14325
 FT TRANSMEM 14345 14365
 FT TRANSMEM 14385 14405
 FT TRANSMEM 14425 14445
 FT TRANSMEM 14465 14485
 FT TRANSMEM 14505 14525
 FT TRANSMEM 14545 14565
 FT TRANSMEM 14585 14605
 FT TRANSMEM 14625 14645
 FT TRANSMEM 14665 14685
 FT TRANSMEM 14705 14725
 FT TRANSMEM 14745 14765
 FT TRANSMEM 14785 14805
 FT TRANSMEM 14825 14845
 FT TRANSMEM 14865 14885
 FT TRANSMEM 14905 14925
 FT TRANSMEM 14945 14965
 FT TRANSMEM 14985 15005
 FT TRANSMEM 15025 15045
 FT TRANSMEM 15065 15085
 FT TRANSMEM 15105 15125
 FT TRANSMEM 15145 15165
 FT TRANSMEM 15185 15205
 FT TRANSMEM 15225 15245
 FT TRANSMEM 15265 15285
 FT TRANSMEM 15305 15325
 FT TRANSMEM 15345 15365
 FT TRANSMEM 15385 15405
 FT TRANSMEM 15425 15445
 FT TRANSMEM 15465 15485
 FT TRANSMEM 15505 15525
 FT TRANSMEM 15545 15565
 FT TRANSMEM 15585 15605
 FT TRANSMEM 15625 15645
 FT TRANSMEM 15665 15685
 FT TRANSMEM 15705 15725
 FT TRANSMEM 15745 15765
 FT TRANSMEM 15785 15805
 FT TRANSMEM 15825 15845
 FT TRANSMEM 15865 15885
 FT TRANSMEM 15905 15925
 FT TRANSMEM 15945 15965
 FT TRANSMEM 15985 16005
 FT TRANSMEM 16025 16045
 FT TRANSMEM 16065 16085
 FT TRANSMEM 16105 16125
 FT TRANSMEM 16145 16165
 FT TRANSMEM 16185 16205
 FT TRANSMEM 16225 16245
 FT TRANSMEM 16265 16285
 FT TRANSMEM 16305 16325
 FT TRANSMEM 16345 16365
 FT TRANSMEM 16385 16405
 FT TRANSMEM 16425 16445
 FT TRANSMEM 16465 16485
 FT TRANSMEM 16505 16525
 FT TRANSMEM 16545 16565
 FT TRANSMEM 16585 16605
 FT TRANSMEM 16625 16645
 FT TRANSMEM 16665 16685
 FT TRANSMEM 16705 16725
 FT TRANSMEM 16745 16765
 FT TRANSMEM 16785 16805
 FT TRANSMEM 16825 16845
 FT TRANSMEM 16865 16885
 FT TRANSMEM 16905 16925
 FT TRANSMEM 16945 16965
 FT TRANSMEM 16985 17005
 FT TRANSMEM 17025 17045
 FT TRANSMEM 17065 17085
 FT TRANSMEM 17105 17125
 FT TRANSMEM 17145 17165
 FT TRANSMEM 17185 17205
 FT TRANSMEM 17225 17245
 FT TRANSMEM 17265 17285
 FT TRANSMEM 17305 17325
 FT TRANSMEM 17345 17365
 FT TRANSMEM 17385 17405
 FT TRANSMEM 17425 17445
 FT TRANSMEM 17465 17485
 FT TRANSMEM 17505 17525
 FT TRANSMEM 17545 17565
 FT TRANSMEM 17585 17605
 FT TRANSMEM 17625 17645
 FT TRANSMEM 17665 17685
 FT TRANSMEM 17705 17725
 FT TRANSMEM 17745 17765
 FT TRANSMEM 17785 17805
 FT TRANSMEM 17825 17845
 FT TRANSMEM 17865 17885
 FT TRANSMEM 17905 17925
 FT TRANSMEM 17945 17965
 FT TRANSMEM 17985 18005
 FT TRANSMEM 18025 18045
 FT TRANSMEM 18065 18085
 FT TRANSMEM 18105 18125
 FT TRANSMEM 18145 18165
 FT TRANSMEM 18185 18205
 FT TRANSMEM 18225 18245
 FT TRANSMEM 18265 18285
 FT TRANSMEM 18305 18325
 FT TRANSMEM 18345 18365
 FT TRANSMEM 18385 18405
 FT TRANSMEM 18425 18445
 FT TRANSMEM 18465 18485
 FT TRANSMEM 18505 18525
 FT TRANSMEM 18545 18565
 FT TRANSMEM 18585 18605
 FT TRANSMEM 18625 18645
 FT TRANSMEM 18665 18685
 FT TRANSMEM 18705 18725
 FT TRANSMEM 18745 18765
 FT TRANSMEM 18785 18805
 FT TRANSMEM 18825 18845
 FT TRANSMEM 18865 18885
 FT TRANSMEM 18905 18925
 FT TRANSMEM 18945 18965
 FT TRANSMEM 18985 19005
 FT TRANSMEM 19025 19045
 FT TRANSMEM 19065 19085
 FT TRANSMEM 19105 19125
 FT TRANSMEM 19145 19165
 FT TRANSMEM 19185 19205
 FT TRANSMEM 19225 19245
 FT TRANSMEM 19265 19285
 FT TRANSMEM 19305 19325
 FT TRANSMEM 19345 19365
 FT TRANSMEM 19385 19405
 FT TRANSMEM 19425 19445
 FT TRANSMEM 19465 19485
 FT TRANSMEM 19505 19525
 FT TRANSMEM 19545 19565
 FT TRANSMEM 19585 19605
 FT TRANSMEM 19625 19645
 FT TRANSMEM 19665 19685
 FT TRANSMEM 19705 19725
 FT TRANSMEM 19745 19765
 FT TRANSMEM 19785 19805
 FT TRANSMEM 19825 19845
 FT TRANSMEM 19865 19885
 FT TRANSMEM 19905 19925
 FT TRANSMEM 19945 19965
 FT TRANSMEM 19985 20005
 FT TRANSMEM 20025 20045
 FT TRANSMEM 20065 20085
 FT TRANSMEM 20105 20125
 FT TRANSMEM 20145 20165
 FT TRANSMEM 20185 20205
 FT TRANSMEM 20225 20245
 FT TRANSMEM 20265 20285
 FT TRANSMEM 20305 20325
 FT TRANSMEM 20345 20365
 FT TRANSMEM 20385 20405
 FT TRANSMEM 20425 20445
 FT TRANSMEM 20465 20485
 FT TRANSMEM 20505 20525
 FT TRANSMEM 20545 20565
 FT TRANSMEM 20585 20605
 FT TRANSMEM 20625 20645
 FT TRANSMEM 20665 20685
 FT TRANSMEM 20705 20725
 FT TRANSMEM 20745 20765
 FT TRANSMEM 20785 20805
 FT TRANSMEM 20825 20845
 FT TRANSMEM 20865 20885
 FT TRANSMEM 20905 20925
 FT TRANSMEM 20945 20965
 FT TRANSMEM 20985 21005
 FT TRANSMEM 21025 21045
 FT TRANSMEM 21065 21085


```

Db      365 YDPLDGNELIALMGNTVCGVFRGAFAGSTALSRSVAVESTGKTQIAGLIIAIIIVLY 424
      419 MVLGIYLPKPSVGLALIAVLNKLKSLKQIDTPYILMRKSKIDCCIMVVSFLSFLSL 478
      425 VLAIGLVLPLQSVLAALALGLMKLMQFAIGRLMRKDKYDCLIMTFITFTVLGL 484
      479 PGVAVGVAFSVLVVFQTOFRNGVALAQMVDIYVNPETKYRADIDQIKITTCSP 538
      485 GGLASVAFQQLITVIFRQFPCSTLANIGRNITKNKKDYDMDEPEBVKIFRCPSPI 544
      539 YFANSEIFPKQVY-----AKTVSLQELQO-----DEFNAP 568
      545 YFANIFEFPRKRLIDAVGFSPRLRLRKRNKRLKRIKLOKGLQVTPKGFICTVDPIKDS 604
      569 PTPNNNO-----TPANGTSVSY-ITFSPDSSSPAOSEPPASAEADGEPSDMLASVPPV 622
      605 DEELDNKQLEVLDPINTITDLPFHIMNDL-----PLNTEVRK-----1 644
      623 THTLLIDNSGVSVFDLMGKALAKLSSTYGIKGVFLVNIHAQVYNDISHGVEEDGS 682
      645 SLHSLIDPSAVSFLDVSSVRGLKSLTGEFIRIKVDYIYIGTDDDFLEKLNREYEF-DGE 703
      683 LECKHYFPSTHDAVL 697
      704 VKSSIFEFLTTHDAVL 718

```

RESULT 2

```

A54808
diastrophic dysplasia-associated sulfate transport protein - human
C:Species: Homo sapiens (man)
C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 07-May-1999
C:Accession: A54808
R:Haesthacker, J.; de la Chapelle, A.; Mahrani, M.M.; Clines, G.; Reeve-Daly, M.P.; Daly,
der, E.S.
Cell 78, 1073-1087, 1994
A:Title: The diastrophic dysplasia gene encodes a novel sulfate transporter: positional
A:Reference number: A54808; MUID:95007757
A:Accession: A54808
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-739 <HAE>
A:Cross-References: GB:U14528
C:Gene: GDB:DTD; DTOST
A:Gene: GDB:DTD; DTOST
A:Cross-References: GDB:125421; OMIM:222600
A:Map position: 5q32-q33.1
C:Superfamily: sulfate transport protein

```

```

Query Match      25.0%; Score 968; DB 2; Length 739;
Best Local Similarity 31.0%; Pred. No. 1,5e-66;
Matches 225; Conservative 139; Mismatches 285; Indels 76; Gaps 11;

```

```

      3 QRRPRVVDRAVSLTLPDEFKKDRTYP--VGKRLNNAFRCSSAKIRAVFGLLPVLS 60
      49 RYHNRILIRQ-----EKSDTNKEFEYIKLKQKCCSPAKAKMIMIGFLVLO 97
      61 WLPRYIKRYIIPDLGLSGSGSIQVPOGMAFALLANLAVNGLYSFPLLTFFLPGV 120
      98 WLPRKIDKLNILGDVSGSLVGLVLPQSIAYSLAGQEPVYGLYSPFASIIYFLLGNS 157
      121 HQMVPTGVVISILVGNIC---LQ-----LAPESKFOVNNMTNESYVDTAAM 165
      158 RHISVGFVFLCLMIGETVDRLOKAGYDANASAPSLGAVNSGSLTNHTSDICKSCSY 217
      166 EBERLHVSATLACLVATIQMGLGFMQGFVAIYLSSTIRGEFTWAGLQILISLVLYIFG 225
      218 ---AIWVGSTVFVAGVYVAMGFQVSVYLLSDALLSGFTVGSFTILTSQAKYLLG 274
      226 LTIPTGTGSGIVFTFIDICKNLPHNTNIALSLFALISGAFVLVYKLNARVMKIRFPPI 285

```

```

Db      275 LNLPRNGVSLITWTTHVFRNIRKTNLCDLITSLCLLVLTPTKELNHFKSKLAPAP 334
      286 TEMIVVAVATATISGGCKMRKYHMOIVGEIQRGEPPTPVSVSOMK---DMIGTAFSLAI 342
      335 TELVVVAATLASHFCKLHENTNINSLAGHLPFGF---MPKVEEMULISVAVDAITSI 391
      343 VSYVIMLAKRGLTANKRGYDVDSNOEMIALGCSNFGSFEKIHVICALSVTLAVDAGG 402
      392 IGFALIVSLISEMFAKKHGTAVKANQEVYALGFCNIIIPSFHCFTTSALAKTLVKESTGC 451
      403 KSOVASLCSLVYVITMLVLGILYLPKPSVGLALIAVLNKLKQIDTPYILMRKSKID 462
      452 HTQLSGVWVALVLLVLLVLAFLPFLYSLOKSVLGLVITIVNLRGLRFRDLPKMSISIRD 511
      463 CCIMVVSFLSPFLSLPYGAVGVAFSVLVVFQTOFRNGVALAQMVDIYVNPPTYNR 522
      512 TVIMPTMILSALLSTEIGLVGCSICVILIRTKPKSSLLGLVESEFVSATYKN 571
      523 AADIQIKITTCSPLYFANSEIFRQKVIKTVSLQELQODEFENAPPTDPNNQTPANGT 582
      572 LQTKPGIKIFRFVAPLYINKECFSALYKQTVNPILIKVAMKAKRKIKERVVTLGGI 631
      583 SVSYITFSPDSSSPAOSEPPASAEADGEPSDMLASVPPVTFHTLLIDNSGVSVFDLMGI 642
      632 Q-----DEMVSQLSHP-----LEHTIVIDCSALQFEDPTAGI 664
      643 KALAKLSSTYGIKGVFLVNIHAQVYNDISHGVEEDGSLECKH-----VFPSTHDAVL 697
      665 HTLKEVRRDYEALGIVGLAQCNPYRDSLTINDEY-----CKKEENLTFYSVEAMA 717
      698 FAQAN 702
      718 FAEVS 722

```

RESULT 3

```

A49994
sulfate transporter, canalicular - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 20-Jan-1995 #sequence_revision 20-Jan-1995 #text_change 24-Sep-1999
C:Accession: A49994
R:Bissig, M.; Hagenbuch, B.; Stieger, B.; Koller, T.; Meier, P.J.
J. Biol. Chem. 269, 3017-3021, 1994
A:Title: Functional expression cloning of the canalicular sulfate transporter system of
A:Reference number: A49994; MUID:94132077
A:Accession: A49994
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-703 <BIS>
A:Cross-References: GB:L23413; NID:9431452; PIDN:AA17545.1; PID:9431453
C:Superfamily: sulfate transport protein
C:Keywords: glycoprotein; transmembrane protein

```

```

Query Match      22.0%; Score 853; DB 2; Length 703;
Best Local Similarity 29.3%; Pred. No. 1e-57;
Matches 201; Conservative 143; Mismatches 304; Indels 38; Gaps 8;

```

```

      36 KLRNAFRCSSAKIRAVFGLLPVLSWLPKYKIDYIIPDLGLSGSGSIQVPOGMAFALL 95
      34 RLKKSCTGSPCAQALVQGLFPIRWLPQRLKEVYLAGVMSGLYGIILVPOALVSL 93
      96 ANLPAVNGLYSFEPLLTFFFLGVQWMPGTFVVISILVGNIC---LQLA----- 143
      94 AGIQLPSLYLTSFPAFLIFLMGTSNHWVGFSLCLWVGQVVDRELQADPDPSQDSL 153
      144 -PESKFOVENN-ATNESYVDTAAMEERLHVSATLACTATITQMGIGFMQGFVAIYLS 201
      154 GPCNNDSITNNTATLVGQDGRDCHAIRATATLTMAGLVQVLMGILRLGFVSTYLSQ 213
      202 SFIRGFMTAAGQIILSVLKYIFGLTIPSYTPGSGIVFPEFIDICKNLPHNTNIALSL 261
      214 PLLDGFAMGASVITLTSQAKHLLGVRIPRHOGIGAVIHTWLSLQANVGANLQDVVTS 273

```

[illegible]

```

RESULT      4
T27820
hypothetical protein zk287.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #ext_change 04-Mar-2000
C:Accession: T27820
R:McMurray, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z20425
A:Accession: T27820
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-611 <MIL>
A:Cross-references: EMBL:Z70757; PIDN:CAA94798.1; GSPDB:GN00023; CESP:ZK287.2
A:Experimental source: clone ZK287
C:Genetics:
A:Gene: CESP:ZK287.2
A:Map position: 5
A:Introns: 3/3; 102/1; 438/3
C:Superfamily: sulfate transport protein

Query Match      20.2%; Score 781; DB 2; length 611;
Best Local Similarity 28.3%; Pred. NO. 3e-52;
Matches 197; Conservative 136; Mismatches 262; Indels 98; Gaps 1

QY      18  TLFDDEEKKDRTYPVEGKELRNAFRCSSARIKAVVFGLLPVLSTWLPKRIKDYIIPDLG 77
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      9  TTTQDQIVKEKCECPFRKTIQK-----YIIIEKLPRYQMKKDHFGHGVIA 53

QY      78  GLSGSGSIQVPOGMAFALLANIPAVNGILYSFFLLTYFFPLGCVHOMVPGFFAVISILVGN 137
      || : : ||||| : || : : ||||| : || : : ||||| : || : : ||||| : ||
Db      54  GLTGVIMIVPOGMAVYASLAGVPVYGYMYSSEFASTIYMFEGTARHISIGVFAVASMVGA 113

QY      138  ICLDAPESKQVQNNATNESYVDFTAMAEARLHVSAFLACITAIOMGSGPMQGFVAI 197
      | : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      114  AKRLRAPD--IPINSSQINSPVPLCEIYVDPLFTSALTLLGCVQIIMGLILKGLFTT 171

QY      198  YLSEFIRNGEFTWTAAGQILISVLKTYIFGTLIPSTYTGSGSIVFTFIDICKMLPHTINASLI 257

```

Db	172	YLSLSLVSAGTTGAANHVFISLQNKAFGKIKLRHHEIGIAVMYKDMIMSLMSFNVALG	231
QY	258	FALISGAFVLVYKELNARYHKI-----RPPIPEMIYVVVTAISGCKKPKRYHMIV	312
Db	232	ISI----FGILFDLDTGTYINPIVKKFESPIPLLELILVFIGIVISMIFINDAEYHKT	287
QY	313	GEIQRGPPPTVSPVYSQMKMIGITAKSLAIVSYVINLARGRLAKKHGVDVDSNQEMAL	372
Db	288	YEIPRGPDLSIPRLNPLPALLSDALPIAVVCMFVMSGKLEAKKHRYKTDAAQOELAI	347
QY	373	GCSNFESFPFKIHVTCALSVLTVAVGAGGKSOVASLCSLVMMITMLVLGIYIPLPKS	432
Db	348	GIALSLSEFPFVYPVCGASLSRSSVCEMSGANGLYTIFFSEFLILVILLLGPELPEPMC	407
QY	433	VLGALIAVNLKNSLKQLTDBEYVLMRKSKLDCCIWVVSFLSEFSLPYGVAVGAESVLV	492
Db	408	ILACIVIVSKLSLFMOVKELPRLYRISKYDFALIMLACASTFTFTVTGVLVSLAFSLYT	467
QY	493	VVPQTORNGVLAQAVMDIDVYNPKTYNRADI--GKIITIIYCGPILFANSEIPIROKVI	551
Db	468	LVLKQOMP-----SSTVYLHDE--TPR-----QVPENAVIVAFAGSLPHANTVA-----	511
QY	552	AKTYSLOELQODEENAPPTDPNNNQTPANGTSVSYITFSPDSSSPAOSSEPPASAEAPGP	611
Db	512	-----QDDM-----GEA	518
QY	612	SDMLASVPPVYTFHTLLLDMSGVSFVDLAKIKALASTYKIGVKYFLVNIHAQVND	671
Db	519	IGKLPREDPLIDERTIILDDASVAFLDIMGVDAIRDVFDALKLGVHVEYCGLPEDVLSV	578
QY	672	ISHGVFEDDSLCKHYFSPSIDHVALEAONARDY	706
Db	579	LSNDENFL-SVWPPSTTFPSI--DSCLISYHQOSV	611

```

RESULT      5
723628
hypothetical protein K12G11.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T23628
R:Kershaw, J.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19774
A:Accession: T23628
A>Status: preliminary; translated from GB/EMBL/DDBU
A:Molecule type: DNA
A:Residues: 1-749 <WIL>
A:Cross-references: EMBL:Z81570; PIDN:CAB04606.1; GSPDB:GN00023; CESP:K12G11.1
A:Experimental source: clone K12G11
C:Genetics:
A:Gene: CESP:K12G11.1
A:Map position: 5
A:Introns: 48/3; 128/1; 165/1; 220/3; 407/3; 471/3; 503/3; 539/2; 585/2; 629/3; 669/3
C:Superfamily: sulfate transport protein

Query Match          20.1%; Score 776.5; DB 2; Length 749;
Best Local Similarity 28.2%; Pred. No. 8.9e-52;
Matches 206; Conservative 142; Mismatches 311; Indels 71; Gaps 13;

QY      2 SGPRRRYV--VDRAYSLTLPDDEFKKDRYPVGEXLRMAFRCSASAK----- 47
       |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      31 SEPNNRRVSFVORGMMNOAQDEKF--DYKNP---HLENELKQAKKFFRRFEPTTSF 84

QY      48 --IKAVNGGLPVLSMLRPFKIKDIYIIPDLGGSGSIYPOGMAFALLNPVANGLY 105
       ::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      85 FALKLFIFDLPIFLMKPFPEYKKKTDLSDIIIGITIVGVQMVPQGIALVALAKQPAINGLY 144

QY      106 SSEFPLLTYFLFGCYHQVMPPGTFAVISILVGNICQLAPESKFQVFNNATNESYVD7TAA 165
       :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      145 TSLEPPLLYMIEGSTRSAHSLSGFVAVVSIMTGSLSVETKLAPTDYD--PSSFNEDIDLVKL 202

```

```

OY 166 EAERLHVSATLACTTAITOMLGFMQFGVAIYLSSESRIGFMTAAGLOILISVLKRYIFG 225
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 203 PSP-TEVSCALITMGLIFWGVLRLOFLTLYTSLDQVYAGTGVSSVHVLVSQKTLTG 261
OY 226 LT-IPSYTGPSSIVFTFIDICKNLPHNTNLSLIFALISAEFLVLYKELNARKMH--KIR 281
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 262 INGLRHSQPFYLRPHLYDLVNSITRANFVSCGISIVSLIIVHFGKEFINPPIKRRKTSN 321
OY 282 FPIPTMIYVVAATAISGCGMKPKRYHMOIYGEIORGFPPTPVASQMKMDIGTAFLSLA 341
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 322 IPIPELIVILSTYFVAVTGVTEAKQVVKIPGVNENLPSFYLLIPVLPDAISIT 381
OY 342 IVSYIYNLMTGTLANKHGYDVSNOEMIALGCSNFECSFEKIHVICALSVTLAVDAG 401
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 382 VYSISVWLSISKMLAKRYNELDSGOELFALSFISSSFPISIPNSCLSRILVAVGAG 441
OY 402 GKSQVASLCSVLYVMITMLYIGIYLYPLPKSVLGALLIANLNKSLKOLDTPYLYRKSKL 461
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 442 CTTOISIFSSSLIVSVVIFLSTLLETLPMAISAIICALOGMRKRFADLIDLKRVSKI 501
OY 462 DCCIMVWSFLSSFFLSPYGAVAVAFSVLVVFOFNGYALAOVMDTDIYVNPRTYN 521
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 502 DPTIVWGVSTLILDVSTGLISCFALFTTLREOYPRKMHLLASVKTODFRDAERYG 561
OY 522 RAQDIQGIKITCYCSPLFANSEIRQKVIATVSLQF-----LQDFENAPPTDPNNN 575
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 562 ETVYFKGICIFRFDAPLLFHNVECFKKSIEKAYTEMOKSHEFYVLRERETILNTKLDGS 621
OY 576 QTPANG-----TSVYI-TSPDSSSPAQSEPPASAEAPGSPDMLASVPPYFTFTLLID 630
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 622 DESIDGKMFQTAOSTLNTNHSPI-----ILSRHEVID 652
OY 631 MSGSVFDLMGKALAKLSTYTGKIGVNVFLVNIHAQVYNDISHGVFEDGSLCKHVP 690
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 653 CSGFTFIDLMGVALKEIFSDMKRKGILLVFANAKAPVREMEKCHFFNFVSK--NFYR 710
OY 691 SIHDAVLEFAQ 700
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 711 TMRDATSIAR 720

```

RESULT 6
 T23629
 hypothetical protein K12G11.2 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
 C:Accession: T23629
 R:Kershaw, J.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z19774
 A:Accession: T23629
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-700 <M11>
 A:Cross-references: EMBL:Z81570; PIDN:CA804607.1; GSPDB:GN00023; CESP:K12G11.2
 A:Experimental source: clone K12G11
 C:Genetics:
 A:Gene: CESP:K12G11.2
 A:Map position: 5
 A:Introns: 26/3; 55/2; 107/1; 197/3; 253/3; 307/3; 335/2; 480/3; 516/2; 562/2; 608/3; 66
 C:Superfamily: sulfate transport protein

Query Match 19.9%; Score 768; DB 2; Length 700;
 Best Local Similarity 28.3%; Pred. No. 3.7e-51;
 Matches 199; Conservative 141; Mismatches 322; Indels 40; Gaps 11;

```

OY 12 RAAYSLTLEDFDEFEKKDRTYVGEKLRNAPRCSS-----AKIKAVFGCLPVLSQL 62
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 21 RAPMNVQVDEKYGQYKRLKDGKFKRSTKVASRYVVPFTSVTFKIFLLNLNLPFGML 80
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 63 PKYIKQVITIDLLGGLSGSSTQVPGMAFALLNLPVAVNGLIYSSFFLLTYFLGCVHQ 122

```

```

Db 81 PRYDMKNLSLTSOVVGITVGVLIQIPQGLAVALLSRODPVGLTYTSIYVFLIFFGTSKH 140
OY 123 NVPGTEFAVSLIVGNICLOLAPESKEOVFNNTNSYVOTAMAEERLHVSATLACTLNAI 182
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 141 ASLGTFAVVALMTG-----LAVREAFIPSDMLNSTLLPGBDALSPLEVSICALVAGCL 195
OY 183 IOMGLFMQFGVAIYLSSESRIGFMTAAGLOILISVLKRYIFGLT-IPSYTGPSSIVFTF 241
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 196 IQFLMGVFLQGLTLYTSLDQVYAGTGVSSVHVLVSQKTLTG 255
OY 242 IDICKNLPHNTNLSLIFALISAEFLVLYKELNARKMH--KIR 298
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 256 YDTVNLPRANMCCASISLATMILLHCGKEYINPIMKRMKSNIPIMELVAVIISTLEV 315
OY 299 GCGMKPKRYHMOIYGEIORGFPPTPVASQMKMDIGTAFLSIAIYSYIYNLMTGRTLNV 358
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 316 ALIDANELYNKVIYKIPGDELSPNPLIPRVPDAISIAVAVVAHLSKMLAKK 375
OY 359 HGYDVSNOEMIALGCSNFECSFEKIHVICALSVTLAVDAGKRSQVASLCSVLYMII 418
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 376 YEYEDAGQELVALSFTALGGSFFPTFTSICLGRTWGVESGVKTQVATFSPCLVLSV 435
OY 419 MLVIGIYLYPLPKSVLGALLIANLNKSLKOLDTPYLYRKSKLDCCIMVWSFLSSFFSL 478
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 436 SLYFGRFLETPMCVLSAIIYALISMLKRLDLGIMKLSIDCCIMVAVFAFVAVLDV 495
OY 479 PYGVAVGAFSVLVVFOTFNGYALAOVMDTDIYVNPRTYNRAQDIQGIKITCYCSPL 538
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 496 SGGLLIAIFALFTTLREOYPRKMHLLASVKTODFRDAERYG 555
OY 539 YFANSEIRQKVIATVSLQELQDFENAPPTDPNNNOPTANGTSVSYITSPDSSSPAQ 598
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 556 LFHNVECFK-KCIEKYVDEMKKSEF-----NFWEKPAKCGSKFTEFGMRIRAPI 605
OY 599 SEPPASAEAPGSPDMLASVPPYFTFTLLIDMSGVFDLMGKALAKLSTYTGKIGV 658
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 606 TEIPLH---PGINRD-----PILPRH-FVIDCSGFTFIDLMGVALKEVSDLRKRRVQ 655
OY 659 VFLVNIHAQVYNDISHGVFEDGSLCKHVPFSDHDAVLEFAQ 700
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 656 VYFASTKVPVREMEKCSFFDFVSK--NFYTLTLDATGIAR 695

```

RESULT 7
 T16077
 hypothetical protein F14D12.5 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
 C:Accession: T16077
 R:Minx, P.
 submitted to the EMBL Data Library, November 1995
 A:Description: The sequence of C. elegans cosmid F14D12.
 A:Reference number: Z18457
 A:Accession: T16077
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-650 <MIN>
 A:Cross-references: EMBL:U41021; NID:g1086679; PID:g1086682; PIDN:AAA82335.1; CESP:F1
 C:Genetics:
 A:Gene: CESP:F14D12.5
 A:Introns: 3/3; 53/2; 77/1; 159/3; 181/2; 215/3; 270/3; 321/1; 443/3; 479/2; 503/3; 5
 C:Superfamily: sulfate transport protein

Query Match 16.3%; Score 631; DB 2; Length 650;
 Best Local Similarity 24.1%; Pred. No. 1.2e-40;
 Matches 179; Conservative 145; Mismatches 290; Indels 128; Gaps 17;

```

OY 25 EKKDRTVP-----GEEKLRNAPRC-----SSAKIKAVFGCLPVLSQLPKYIKQVYI 72
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4 EBYDMYFNFNDEGNETVWKKNIN--YCTPLSKDGIKRIILRRVPIDMIGSYQIINN-A 60

```

QY 73 PDLGLGSGGSIQVPOGAAFLANLPAVNLGYSFPELLTYFFLGAGVHQVWPTCPAVIS 132
 Db 61 SDFIAGTLMGYVNPQAMSYSLAGLPVYGLYASFPPPELXSLIFGSAKHSIGVFSTC 120
 QY 133 ILVGNICLQ LAPESKFOVENNATNESYVDTAEMERLH-----VSATLACTAII 183
 Db 121 LAMVK-CVK-----KMLFRNENPES-ELISFSDKLYNMAQNSPAVMAIFPCDKMK 172
 QY 184 QMGCGFMOFGVAIYLSSEFIRGFMTAGLOILISVLKITYIGLTIPTGTGC-SIVFPI 242
 Db 173 FLGAPAIIS---ATFESCFEG-----VVSQIPKLGCFVSPRSHMFSLVSHSL 218
 QY 243 DICKLPHPTNASTLIFALISGALF---VLKELNARMKIRREPIEMLVVVAVATAIS 298
 Db 219 DIFECHSNATATLISASALVFLIGSRIFIEPFFKNKKLQSIIPFKELITVIATAS 278
 QY 299 GGCKMKPKRYHMQVGEIQRGFPTPVSPVSOQMDIGTAFSLATVSYVINLMDGRTLANK 358
 Db 279 YEFDFEHFVGKTLHTVRCRGPYDPVPRIDIMPIYIFODALSIAYAVAVTMAQGEFATK 338
 QY 359 HGVDVDSQEMIALGCSNFSSFKINHYICALSVTLAVDAGGSKQVASCVSLVMYIT 418
 Db 339 HRYRDSNQELLALGFINIGSSFVSFPTSAFSRTLYNENSGAKTOLDIGTISACFMALV 398
 QY 419 MLVGLIYLPPLPKSVGLAVNLKNSLKOLDYPIYLRKSKLDOCIWVSFLSFLSL 478
 Db 399 ITTIGPYIASLPSCLSAIYIVLESMLRKYTLPLGRCKSKHDFWITITAVTLSSDI 458
 QY 479 PYGAVGAVSFLVVFQTOGRFNGYALAAQVMDTDIYNPKTYNRAODIGIKITTYCSPL 538
 Db 459 AOGVAGAGIFTAIFTAIGSQOPTIKLQIIRPNDF--RPMNHYSIAKETSPRIIRFAPDL 516
 QY 539 YFANSEIRKQVIKTYVLSLOELQDFENAPPTDRNNOTRANGISVSITTFSPSSSPAQ 598
 Db 517 IFANVEKLDNV-----RSTSAAT 536
 QY 599 SEPPASAEPEPSDMLASVPFVTFTLLIDMSGVSEVDLMGIKALKLSYTGKIVK 658
 Db 537 KKRASCSQSD-----IEMSAIILDCHTWITDMSGIAAAYQINDELROIFIL 585
 QY 659 VFLVNIHAQVYNDISHGGEFEDGSLCKHVPESIHDAVLAQANARADYTPGHNFQAGBD 718
 Db 586 LFLFAULKSLRQYGNAGLIEGNGIRNQMPYSIODAL---DSAHLELIKNN----- 633
 QY 719 AELSLYDSEEDIRSYWDLEQEM 740
 Db 634 -----KMFDECKSM 643
 RESULT 8
 T25751
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T25751
 R:Fulton, B.
 Submitted to the EMBL Data Library, August 1996
 A:Description: The sequence of C. elegans cosmid F41D9.
 A:Reference number: Z20081
 A:Accession: T25751
 A:Status: preliminary: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1758 <FNU>
 A:Cross-references: EMBL:U67954; PIDN:AA852608.1; GSPDB:GM00028; CESP:F41D9.5
 A:Experimental source: strain Bristol N2; clone F41D9
 C:Genetics:
 A:Gene: CESP:F41D9.5
 A:Map position: X
 A:Insertions: 5/3; 74/1; 111/1; 165/2; 206/3; 261/3; 313/3; 370/3; 466/3; 502/2; 551/3; 585/3
 C:Superfamily: sulfate transport protein

```

Best Local Similarity    24.6% ; Pred. No. 3, le-37;
Matches 175; Conservative 141; Mismatches 295; Indels 100; Gaps 11

Oy      35 EKLBNARFCSAKTAKAVGGLLPVLSMLPKYKTKDYIIDPLGLSGSIOVPQGMAPL  94
       :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db      22 QKLR--YACSPSKCHSLSFLLPITLMPLPYDMSHSFEFGDLSCGLTMFAVSPQGIAS  79

Oy      95 LANIPANGLYSSPEFLTYEFELGGVHQWMPGFVAVISIL----VNICIGOLA-----  143
       :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db      80 ITGVPPVYGLTATLEFPSEFLTYEFTGSKHNAALCGFAVLSTTHCAIEKVMIRTAITSNATA  139

Oy      144 -----PESKFQVFNNAT-----NE-SYVDPAAME-----AERLHASATLA  177
       :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db      140 YVNHTELDDELDKENETALISNTLMQTLGNETSFEVEEMMTEGVTPKQIHVATFTII  199

Oy      178 CLTAIIOGCGFMQPGVAITYLSSFLRGTMTAAQDILISLYKTIFFGLTIPSYTPRGSI  237
       :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db      200 FLAVAIQVPMKVLFQLQYLTSIFSEQVAGSVGGGHVFFAQIGNMIGIELPRSGYGL  259

Oy      238 VFTEFDICKNLPHNIINSILFALISGAFLVLELNARYMHR--IRPIPTEMIIVAVARA  296
       :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db      260 YYRIIMDLVENIDNNHIPTVCISLSSFLEVFGKEYLAMPINLSMFNPVPRELVILL-----  315

Oy      297 ISGCCMKPKKYHMOIVEIORGEPTPVSPVVSQMKDMIGAFSLAIVSYVINIAMGRTLA  356
       :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db      316 -----HYLI-----NRFPPELPREDLIRHIGLMAAAIAITVAVAIHITVAKVE  359

Oy      357 NKHGVDYSNOEMIALGCNSNFGEFFIHHICALYSTLAVDAGAKSOVASICVSLAYM  416
       :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db      360 KRYIKKINHGOELTALDEFGVLSSEFFVFPVTSGFANSVGAAVGSGTOLTCIFFSSLAIL  419

Oy      417 ITMLVLGIYLPKRKSVALGALIAVNLKNSLKOLDTPYYLRMRKSLDCDIWVSSFLSFFL  476
       :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db      420 SVILCIGPALEYLPQCILSAMITIPAOKGMLEKEGELKSLMPVKRIDPTIWLMSFLTVCY  479

Oy      477 SLTPGVAVGAFSVLVVVFOQPFNGALAAQMDTDIYVMPKRYNNRADQG-IKIITYC  535
       :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db      480 DMGEGLLMAIGFAVLTITITFORKMHFLSRDDTFENYKETKKRD-LERIQGWVICFRMD  538

Oy      536 SPLFANSSEIRKQVIKTAVSLQLODFENAPPTDENNNQTANGTSVSYIR---FSPD  592
       :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db      539 APLIFTSSDR-----TMSVQCCKMKMERC-----KSESEFIYIEQMND  577

Oy      593 SSSPAQSEPPASAE---APEPSMLASVPPYTFHTLLIDMGVSFVLDLMIKALAKUS  649
       :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db      578 RSADIFDSKLKSARRRRKROQKSENRC-----LVIDCGFPFYVDLGLSTLSYV  628

Oy      650 STYCIKGKVFVLNIIHQVYNDIHGCVFEDGSLECKHVPSTIHDVLAFLAO 700
       :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db      629 VDLQAAGIGCTCFV-VQKSDLKRLRAIDTFEYVDESK-VENKGDVAKAE 677

RESULT          9
T26165
hypothetical protein W04G3_6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-oct-1999 #text_change 18-Feb-2000
C:Accession: T26165
R:Renard, N.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z20163
A:Accession: T26165
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-380 <MW>
A:Cross-references: EMBL:Z68014; PIDB:CAA92028.1; GSPDB:GN00028; CESP:W04G3_6
C:Genetic source: Clone W04G3
A:Comments:
A:Map position: X
A:Introns: 46/1; 83/1; 135/3; 153/3; 190/3; 244/3; 272/2

```

Query Match 14.3%; Score 554.5; DB 2; Length 380;
Best Local Similarity 33.6%; Pred. No. 4,6e-35;
Matches 119; Conservative 78; Mismatches 148; Indels 9; Gaps 3;

Dy 56 LPLVSLWPKYKIDYITIPDLGLSGSIOVPQGMARFALLANIPAVNGLSFFPLLTYF 115
: : | | : : : : : : | | | | : : | | : : | | : :
Db 13 IPIIKLPMYSISENLINDIGVTGILNVPQMAASLVGLPVGLYTSLPSLIYM 72

Oy 116 FLGGVHQMVPGFPAVISIILVGNICLOLAPESKPFQVENMATNESYVDPAAMEERLHVSA 175
| : | : | : | : | : | : | : | : | : | : | : | :
Db 73 FGTSHRVALGVFAVVSLMSGSCNLRKYOEI--ISNNGN---MTSEVEVAISINVK 126

Oy 176 LACLTAIOMGLCFMOPGFPAIATLSSEFINGEFTMAAGLIIVSLKTIIFGLTIPSYPG 235
: : | :
Db 127 LCLAVGLIOLIMGLFRANYDISIIDOIIIGFTGAAVHLTLADLNKLILGVALPRHSIG 186

Oy 236 STVFPEFDICKMLPHNINSLIFALISGAFVLVKE--LNARMHKIRPIEMTVV 292
: : | : | : | : | : | : | : | : | : | : | : | :
Db 187 KLFETIYODLTITAIIGKRVNTITGASVAIVILIISKYILTTPRCATLPVIPEDFLIV 246

Oy 293 VATAISGCCMKMPKHMQIAGEIORGFPPVPSPVSQMKDMIGAFLAIVSYINLAM 352
: : | : : : : : : : : : : | | : : | : : : : : : :
Db 247 ICTIADMFOVERERMGVKVGDLPTEGPFPNALDLEFRHIVIGALAIATSVVYTSMG 306

Oy 353 RFLANKRGYDVDSNQEMIALGCSNFEGSEFKIHVICALSVTIAVDAGCKSQY 406
: : | : | : | : | : | : | : | : | : | : | : | :
Db 307 KYIAWKHNIEIDYROEFPAIGIVASTCSMPCPATALTARTLINDNAGTKYQ 360

RESULT 10
B86365
Probable sulphate transporter protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
Accession: B86365
R:Theologis A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hulzar, L.
Nature 408, 816-820, 2000
C:A.: Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luo, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence number and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: AB6141; MUID:21016719
A:Accession: B86365
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-631 <STO>
A:Cross-references: GB:AE005172; MID:g2829902; PIDN:AAC00610.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1
C:Superfamily: sulfate transport protein

Query Match 14.2%; Score 549; DB 2; Length 631;
Best Local Similarity 26.0%; Pred. No. 2.5e-34;
Matches 175; Conservative 120; Mismatches 251; Indels 126; Gaps 20;

Dy 18 TLDFDEFEKKDTTPYGEKRLRNARCCSSAKIKAVNGELPVLWSLPRYKTKDYIIPDLG 77
: : | : | : | : | : | : | : | : | : | : | : | :
Db 25 TFFPDPLPLQFGOPRKRL-----IRAAY-IPIIIONMCEPYSF-SLLKSQVS 72

Oy 78 GLSGSIOVPQGMARFALLANIPAVNGLSFFPLLTYFLLGVGHQMPGFPAVISILVN 137
| : | : | : | : | : | : | : | : | : | : | : | :
Db 73 GLTISLAPQGISIAKLNLNPVIGLVSEFPPLVAVYAVLGSSRDLAIVGPSIASILLGS 132

Oy 138 ICILAPESKPFQVFNATNESYVDPAAMEERLHVSAATLACTAIITAIOMGLCFMOPGVAI 197
: : | : | : | : | : | : | : | : | : | : | : | :
Db 133 MLRQ-----QV-----SPYDDPVLFQ-LAFESTF--FAGLVOASGLIRLGFITD 175

Oy 198 VLSESEFINGEFTMAAGLIIVSLKTIIFGLTIPSYPGSIPTFDICKMLPHTN---- 252

[illegible]

QY 173 SATLACTLAIIITOMGFGFQNFQFVAIYLSSEFSGFRGFWTAAGDIIILSVKTYEGLTIPSYT 232
Db 167 AFTAFTEFGVLEASLGITRIGFYDFELSHATITVGMGAATVYSIQDQKGFGL--KFTT 224
QY 233 GPGSIVFTFIDICKNLPHTNIASLIFALISGAFVLVYKELNARMYHKIRPPIEMIVV 292
Db 225 DSTDIYSVMRSVFSTHEMRWES---GVLCGCFLPFL--LSTRY-----FSIKRKPFW 274
QY 293 VATA-----ISGGCM-----PKYIMQIVGELQRFPPRY-----SPVSSQMKDIG 335
Db 275 AAMPPLTIVILGSLIVFTTHAERHGVQVIGDLKGL-NPLSGSDLIFTSPYMS--TAVK 330
QY 336 TAFSLAIVSYVINLAMGRTLANKHGVDVDSNOEMIATLCSNFFGSEFFKIHVICALSVTL 395
Db 331 TGLITGIIALABGIAVGRSPFAMFKYNIIDCKNEMIAFEMNIVGSFTSCYLTITGFFSRSA 390
QY 396 AVDAGAGKSQVASLCSVLNMTMLVGLIYVPLPKSVGALIAVNLKNSLKQLTDPYLL 455
Db 391 VVNYNGCCTATMSNIYMAIAVAVFTLLFTLPFLPHYPLVIVLSAII--ISAMGLIDYQAAIHL 449
QY 456 WRKSLDLCICIMVVSFLSFPFLSPYGAVAVGAFSVLVVVPQTFQFNGALNAQVMQTFDIY 515
Db 450 WKVDKFDPLVCKSAIVGVYGVSGVEIGLVAAVAIAISTARLLFVSRPKTVAKGNIPSMITR 509
QY 516 NPKTYNRAADIQIKIITYCSPLFYANSEIFROKVIYAKTVSLQELQODFENAPPTDPNNN 575
Db 510 NTEQYPSRRTPVGLILEIDAPIYFANASYLERITRM-----IDEEE 553
QY 576 QTPANG--TSVSYITFTSPSSSPQOSEPPASAEAPQEPSDMLASVPFTEHTLLIDMSGV 634
Db 554 RYKSGESSLOYI-----IIDMSAV 573
QY 635 SFVDLMGKAKLAKLSSTYGKIGKVFVLNIHQAQVNDISHGCFEDGSLCKHVPSPSID 694
Db 574 GNIDISGISMVWEIKKYIDRALKLVLSNPKEGVYKKILTRSKFIQD-HLCKEMMLTVIGE 632
QY 695 AV 696
Db 633 AV 634

RESULT 12
T49069
sulfate transporter (ATST1) - Arabidopsis thaliana
N:Alternate names: protein F4F15.10
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 21-Jul-2000
C:Accession: T49069
R:Alcazar, J.P.; Clabault, G.; Cottet, A.; Mache, R.; Mewes, H.W.; Rudd, S.; Lemcke, K.;
submitted to the Protein Sequence Database, April 2000
A:Reference number: 225015
A:Accession: T49069
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-658 <ALC>
A:Cross-references: EMBL:AL049711; GSPDB:GN00061; ATSP:F4F15.10
A:Experimental source: cultivar Columbia; BAC clone F4F15
C:Genetics:
A:Gene: ATSP:F4F15.10
Map position: 3
A:Introns: 116/1; 185/2; 242/3; 264/3; 302/3; 341/3; 384/1; 479/3; 543/2; 572/1; 597/3
C:Superfamily: sulfate transport protein

Query Match	14.0%	Score 543.5	DB 2	Length 658
Best Local Similarity	24.9%	Pred. NO. 7e-34		
Matches 180	Conservative 127	Mismatches 274	Indels 141	Gaps 23

```

0y      4 PRPRVVDRAAYSL-TLF-DDEEKDKDRITYPVBEKLRNRCSSAIKAVFEL---LP 57
      1111111111111111111111111111111111111111111111111111111
db      25 PQPQFLKSLQYSVKETLFPPDPFRQ-----FKQNMNR-----KVLGKTFLEP 69
      1111111111111111111111111111111111111111111111111111111
0y      58 VLSQLPKYKIKIDIIIPDLGSLGSGSIQVPGQMAFALLANLPAVNGLYSSFFLLTYFFL 117

```

```

Db      70 IFENAPRNLK -FEKSDIDAGITTA SLAIPGISAKLANLPILIGLSSVPLVYAVL 128
QY      118 GGVHOMVGTFAVISILVGNICLOLAPESKFOVENNATNESYVDTAAEAER----LHV 172
Db      129 GSSRDLAVGVAVASLITG-----AMLskE-----VDAREDKPLYLHL 166
QY      173 SATLAICLTAIIOMGEGFNOGFVAIYIASESPFIRGFMTAAGLIILSVUKIIFGLTIPSYT 232
Db      167 AFTATFAGVLEASIGIRLGFIDFSLHATLVGPGGAATVVSILQIJKIFGL--KHFT 224
QY      233 GPGSIVFFIFIDICKLPHNTINASLIFALISGAFVLVYKELNARVYMHKIRPDIPTEMIYVV 292
Db      225 DSTDIVISMRSVESQTHMRWES---GYLGGCFLFFL--LSTRY-----FSIKRKPFWV 274
QY      293 VATA-----ISGGCKM----PKYHMOIVGETIQRGEPYV-----SPVVSQKMDIG 335
Db      275 AAMAPLTSVILIGSLVYFTHAERHGVQVIGDKLGL-NPLSGSDLIETFSPYMS--TAVK 330
QY      336 TAFSLAIVSYVINIAMGFTLANKHGVDVDSNOEMALGCSNPFSGSEFRIHYICCALSTL 395
Db      331 TGLITGIIALBGAVAVGSPAFMKFNYNIDGKEMIAFMMINIVGSEFCTYLTPGFESKSA 390
QY      396 AVDGAGKSOVASLDCSVLVMITMLVLGILYLPKSVGLALIANLKNLSIKOLDPYLL 455
Db      391 VYNNAGCMTANSNIYMAIAVMFTLLFLPLRPHYPLVYLSAI--ISAMIGLIDYQAAIHL 449
QY      456 WRKSKLDCCIMVVSFLSEFPLSLPYGAVGAFAVSLLVYVFOTOPRNGYALAAQVMDTIYV 515
Db      450 WKVDFEDFLVCMASAVGVGFSVEIGLVAAVAISIAIRLLEFVSREPKTVAKGNIPSMIYR 509
QY      516 NPKTYNRRADLOGIGIITIIYVCSPLYANSEIRPROKVIATVSLQELQOQFENAPPTDPNN 575
Db      510 NTEOYPPSRTYVGLILILIDAPIFANASYLERITIR-----IDEEEE 553
QY      576 QTPANG-TSVSYITFPSPDSSSPAQSEPPASAEAPGSPSMDLASVPFVTFHTLLIDMSGV 634
Db      554 RVKGGESSLOYI-----LIDMSAV 573
QY      635 SFVDIMGIKALAKLSITYGKIGKVFLVNIHAQVYNDISHGCVFEDGSLCKNHVPSIHD 694
Db      574 GNIDISGISMVMEIKKVIDRALKLVLSNPGRGEVYKKLTRSKFIQD-HLKGEMFPLTIGE 632
QY      695 AV 696
Db      633 AV 634

```

```

RESULT 13
T51161
hypothetical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 28-Jul-2000 #sequence_revision 02-Sep-2000
C:Accession: T51161
R:Comella, P.; Wu, H.-J.; Laidie, M.; Berger, C.; Cooke, R.; Delseny, M.; Grellet, F.
Plant MOL. Biol. 41, 687-700, 1999
A:Title: Fine sequence analysis of 60 kb around the Arabidopsis thaliana AtEm1 locus
A:Reference number: 224855, MUID:20108326
A:Accession: T51161
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-703 <COM>
A:Cross-references: EMBL:AF049236; PIDDN:AAC14417.1
C:Genetics:
A:Map position: 3
A:Introns: 161/1; 230/2; 287/3; 309/3; 347/3; 386/3; 429/1; 524/3; 588/2; 617/1; 642/2
C:Superfamily: sulfate transport protein

```

Query Match	14.0%	Score 543.5	DB 2	Length 703
Best Local Similarity	24.9%	Pred. No. 7.7e+34		
Matches 180	Conservative 127	Mismatches 274	Indels 141	Gaps 23

```

OY 4 PRRRYVDRAAYSL--TLF-DDEFEKKDRTYVGEKLRNAPRCSAKIKAVVGL--LP 57
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 70 PQDPFLSKSLQSVKETELEFPDDPFRO-----FKQNASR-----KVLGLKRYFLP 114
OY 58 VLSMLPKYKIKYIIPDLLGLSGSIQVPGMAFALLANPAVNGLYSSFFPLLTYPFL 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 115 IFEMAPRYNLK-FKPSDLIAGITIASLAPGISAKLANLPILGLVSSFFPPLPYAAL 173
OY 118 GGVHOMVGTFAVISILVGNICLOLAPESKFOVFNATNESYVDTAAEAER-----LHV 172
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 174 GSSRLAAGTAVASLLTC-----AMLSKE-----VDAEDPKLYLHL 211
OY 173 SATLACLTATIIOMGLGFMOGFVAIYLSSEIFSGMTAAGLIILSVLKYIFGLTTPSYT 232
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 212 AFTATFFAGVGLASLIGFLGFIVDLSHATIVGEMGAATLVSLQOLKIGFL--KHPT 269
OY 223 GGGSIYVFFIDICKNLPHNTINSLIFALISGAFLVYKELNRYMKIKRFPITENAIYV 292
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 270 DSTDIVSMRSTVPSQTHKRMES---GVLCGGLFEL--LSTRY---FSIKKRFVW 319
OY 293 VATA-----ISGGCKM---PKYHMOIYGEIQRGFPTPV-----SPVVSOMKDMIG 335
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 320 AAMALTVITIGSLVLYFTHAERHGOYIGDLKGL-NPLSGSDLIPTSPYMS---TAVK 375
OY 326 TAFSLATYSYVINLAMGRTLANKHGYDVDSNOEMTALGCSNFFGSEFKIIVICALSVTL 395
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 376 TGLITGIIALAEVAVGVSFAFMKKNYINIDGNKEMIAFGMMNIVGSCYLTTCPEFSRSA 435
OY 396 AVDGAGGSOVASLGVSLVMTMLVLGIYLYPLKPSVIGALIANLKNLSKQLDPYTL 455
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 426 VYNNAGCTTANSIVMAIALMTLELPLPLHYTPLVNLSAIT--ISAMGLIDYQAAHL 494
OY 426 WRKSLDCCIVWVSFLSFFLSPLPGVAVGVAFSVLVVFOTQFRNGYALQAVMDTIYV 515
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 495 WKVDFEFLVCMSAVYGVVGVSGVEIGLVAVAISTARILLFVSRPKTAVKGNIPMSMIR 554
OY 516 NPKTNRAQDIOGKIITYICSPLYFANSEIFRQKIYAKTVSLQELQODEFNAPRPDPNN 575
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 555 NTEQYPSRTYVGLILEIDAPIYFANASYLREKIRW-----IDEEBE 598
OY 576 QTPANG-TSVSYITFSPDSSSPAQSEPPASAPAEPEPSDMLASVPEVTHTLIDMSGV 634
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 599 RKQGESLSQYI-----LIDMSAV 618
OY 635 SFVDMGIKALAKLSSTYVGKIGVKVFLVNIHAQVYNDISHGVFEDGSLCKHVPESIH 694
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 619 GNIDSGISGMVEIKKVIDRALKLVLSNPKEGVVKKILRSKFIDG-HLKGEMFLTVE 677
OY 695 AV 696
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 678 AV 679

RESULT 14
T32945
hypothetical protein W01B11.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Mar-2000
C:Accession: T32945
R:Bradshaw, H.; Graves, T.; Blair, T.
submitted to the EMBL Data Library, January 1998
A:Description: The sequence of C. elegans cosmid W01B11.
A:Reference number: 721250
A:Accession: T32945
A:Status: preliminary; translated from GR/EMBL/DBDU
A:Molecule type: DNA
A:Residues: 1-809 <BR>
A:Cross-references: EMBL:AF043704; PIRN:AAB97595.1; GSPDB:GN00019; CESP:W01B11.2
A:Experimental source: strain Bristol N2; clone W01B11
A:Genetics:
A:Gene: CESP:W01B11.2
A:Map position: 1
A:Insertions: 15/2; 38/3; 114/1; 269/3; 420/1; 497/1; 518/2; 580/1; 613/1; 648/1; 684/1; 72

```

```

C:Superfamily: sulfate transport protein
Query Match 13.8%; Score 532.5; DB 2; Length 809;
Best Local Similarity 23.8%; Pred. No. 6,6e-33;
Matches 178; Conservative 118; Mismatches 274; Indels 177; Gaps 19;

OY 54 GLPLVSLWPKYKIKYIIPDLLGLSGSIQVPGMAFALLANPAVNGLYSSFFPLLT 113
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 79 GRIPITMMLPRTYKSNFLYDPLGLGLMSVLSVPSLSAIGMLVGPSPGLTIGLIPRI 138
OY 114 YFFLGGVHOMVGTFAVISILVGNICLOLAPESKFOV-FNNATNESYVD-----TAA 164
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 139 YALFTSKSHSPGAAIYSLAVGTAV-----ESFGDVGSTNGTIDSNVDLLCCRENKPRV 193
OY 165 MEAEHLASATLACLTATIIOMGLGFMOGFVAIYLSSEIFSGMTAAGLIILSVLKYTE 224
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 194 SDGDAIGTAAVYTLVLVGLLEGLNGLAVMSDOLVGLISGAALVHLSQLSKWT 253
OY 225 GLT-IPSYTPGSIYFTEFIDICKNLPHNTINSLIFALISGAFLVYKELNARYMK---I 280
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 254 RISNVPPISEPPQNVPVFWCFKQMPDYASIIISVLCVLLSTLVIDPWLCKKVPY 313
OY 281 RPIPTFEMIVVVAIISGGCKMPKKYHMQI--VEIQRGFPTPVSPVVSOMKDMIGTAF 338
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 314 KRPME--LVLYVSMTLVHYTRNTAYEFQIPTVGQVTSIGIPAPYIPPRNAFGMGMSAM 370
OY 339 SLATISYVINLAMGRTLANKHGYDVDSNOEMTALG---CSNFFGSEF---KIHVIC-- 388
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 371 STAITSFYHHSCLKLSKLOLYVNSNOEMFALGLMHTTSFSGCFGALDSHHIHCAN 430
OY 389 -----CALSVTLAVDGA-----GKRSQVAS 408
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 431 HRCVWRSCMYSALAEETGARVDCGCCDEGLVSHDFHLRVEIERFCGLRGKPRGYRT 490
OY 409 LCVSLVMTMLVLGIYLYPLKPSVIGALIANLKNLSKQLDPYLYLRKSKLDDCIWV 468
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 491 -----HKSRYANLRCNVHLNRKACQPRPAH-WADSSL----- 523
OY 469 SFLSFFLSPLPGVAVGVAFSVLVVFOTQFRNGYALQAVMDTIYVNPRTYNRADIOG 528
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 524 -----LGRITNTNRRGLGETYANDIPG 547
OY 529 IKIITYCSPLYFANSEIFRQKY-----IAKTVSLQELQODEFNAP 569
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 548 IKIFRDEPIYFGNSEMEFRKIHQACGLNPLIVRGELEETENKKDKARKEEDEAIKPE 607
OY 570 TDPNNQTPRANGTSYITFSPDSSSPAQSEPPASAPAEPEPSDMLASVPEVTHTLIL 629
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 608 TEKE-----TKLEVTHQAKVLDEKHAEPN--PADQYERLTHI-----II 647
OY 630 DMSGVSFVDMGIKALAKLSSTYVGKIGVKVFLVNIHAQVYNDISHGVFEDGSLCKHVF 689
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 648 DCSITIIYDLMQAGLAKDYVNDYKTIIGISVLEFANNENHVRQNFERSQFEE--VPRGNMF 705
OY 690 PSIHDAVLEPAQANADVTPGHNFGAP 716
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 706 VVSDAVDAQAELEQRKTEYQTAAAP 732

RESULT 15
T48902
sulfate transporter AST12 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 21-Jul-2000
C:Accession: T48902
R:Takahashi, H.; Sasakura, N.; Kimura, A.; Watanabe, A.; Saito, K.
Plant Physiol. 121, 686, 1999
A:Title: Identification of two leaf-specific sulfate transporters in Arabidopsis thal
A:Reference number: 225001
A:Accession: T48902
A:Status: preliminary; translated from GR/EMBL/DBDU
A:Molecule type: DNA

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 09:02:14 ; Search time 14.18 Seconds
(without alignments)
1194.991 Million cell updates/sec

Title: US-09-749-589-2

Perfect score: 3869
Sequence: 1 MSQPRPRYVDRRAVSLTLF.....WDLQDFGSMFRAETLAL 753

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCITUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1120	28.9	764	1 US-08-424-567-2	Sequence 2, Appli
2	1120	28.9	764	2 US-08-711-928-2	Sequence 2, Appli
3	1120	28.9	764	4 US-09-184-937-2	Sequence 2, Appli
4	123.5	3.2	470	1 US-08-471-456-2	Sequence 2, Appli
5	123.5	3.2	470	2 US-08-894-840-2	Sequence 2, Appli
6	123.5	3.2	470	3 US-09-139-675-2	Sequence 2, Appli
7	104.5	2.7	886	3 US-09-110-116-3	Sequence 3, Appli
8	103.5	2.7	1956	4 US-08-843-417-10	Sequence 10, Appli
9	103	2.7	438	2 US-08-677-049-9	Sequence 9, Appli
10	103	2.7	834	2 US-08-861-464-4	Sequence 4, Appli
11	103	2.7	834	4 US-09-336-001-4	Sequence 4, Appli
12	103	2.7	834	2 US-09-323-433A-4	Sequence 4, Appli
13	101	2.6	622	2 US-08-132-990A-4	Sequence 4, Appli
14	101	2.6	622	5 PCT-US92-09382-4	Sequence 4, Appli
15	99.5	2.6	462	2 US-08-898-976-2	Sequence 2, Appli
16	99.5	2.6	462	2 US-08-898-976-4	Sequence 4, Appli
17	99	2.6	574	1 US-08-140-729A-7	Sequence 7, Appli
18	99	2.6	574	1 US-08-546-666-7	Sequence 7, Appli
19	99	2.6	574	1 US-08-916-745-7	Sequence 7, Appli
20	99	2.6	574	2 US-08-948-569A-4	Sequence 4, Appli
21	99	2.6	574	2 US-08-663-808-6	Sequence 6, Appli
22	99	2.6	574	2 US-09-042-929-7	Sequence 7, Appli
23	99	2.6	574	2 US-08-546-661-7	Sequence 7, Appli
24	99	2.6	574	2 US-09-042-960-7	Sequence 7, Appli
25	99	2.6	574	2 US-09-188-469-4	Sequence 4, Appli
26	99	2.6	574	3 US-09-198-650-7	Sequence 7, Appli
27	99	2.6	574	3 US-09-332-740-6	Sequence 6, Appli

28	99	2.6	574	3 US-09-042-913-7	Sequence 7, Appli
29	99	2.6	574	3 US-09-188-456-6	Sequence 6, Appli
30	99	2.6	574	3 US-09-042-937-7	Sequence 7, Appli
31	99	2.6	574	4 US-09-397-238A-4	Sequence 4, Appli
32	96.5	2.5	401	3 US-08-492-459-6	Sequence 6, Appli
33	96.5	2.5	401	3 US-08-492-459-8	Sequence 8, Appli
34	96.5	2.5	401	3 US-08-492-459-28	Sequence 28, Appli
35	96.5	2.5	401	3 US-08-492-459-29	Sequence 29, Appli
36	96.5	2.5	401	3 US-08-423-752-6	Sequence 6, Appli
37	96.5	2.5	401	3 US-08-423-752-8	Sequence 8, Appli
38	96.5	2.5	401	4 US-08-716-873-6	Sequence 6, Appli
39	96.5	2.5	401	4 US-08-716-873-20	Sequence 20, Appli
40	96.5	2.5	401	4 US-08-716-873-22	Sequence 22, Appli
41	96.5	2.5	401	4 US-08-716-873-42	Sequence 42, Appli
42	96.5	2.5	401	4 US-08-716-873-43	Sequence 43, Appli
43	96.5	2.5	401	4 US-09-368-431-6	Sequence 6, Appli
44	96.5	2.5	401	4 US-09-368-431-20	Sequence 20, Appli
45	96.5	2.5	401	4 US-09-368-431-22	Sequence 22, Appli

ALIGNMENTS

RESULT 1
US-08-424-567-2
; Sequence 2, Application US/08424567
; Patent No. 5569755
; GENERAL INFORMATION:
; APPLICANT: SCHWEINFEST, Clifford W.
; APPLICANT: PAPAS, Takis S.
; TITLE OF INVENTION: Colon Mucosa Gene Having Down-Regulated
; TITLE OF INVENTION: Expression In Colon Adenomas And Adenocarcinomas
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/424,567
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/026,045
; FILING DATE: 05-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 40399/181 NIHD
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 764 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-424-567-2

Query Match 28.9%; Score 1120; DB 1; Length 764;
Best Local Similarity 34.0%; Pred. No. 1.7e-101;
Matches 250; Conservative 142; Mismatches 275; Indels 68; Gaps 11;
QY 7 RIVVDRAVSLTLFDDFEKKDRTYPVG-EKLRNAFRCSAKIKAVFGLLPVLWMLKRY 65

```

Db      8 QYIVARPVYSTNAFEENHKKTRGHHKTFDLHLKVCSCSPQAKAKIVLSTLFPJASLWPAY 67
QY      66 KIKDYIIPDLGLSGSGSIQVPOGMAFALLANLPAVNGLYSSFFPLTYFFFGVGHQWMP 125
Db      68 RKEWMLSDIYSGISTGIVAVLQGLAFALLVDIPVYGLYASFPAIITLFFGTSRHISV 127
QY      126 GTFVAISILVY----NICQLAPESKFOVE---NNATNESYVDTAAMEARLHVSATLAC 178
Db      128 GFFPLISMVWGLAVGAVSKAVPDRNATTLGLPNNSNSSLDD--ERVRVAAAASYTV 184
QY      179 LTAIOMGLGEMOGFVAIYLSSEFIRGFMTAAGLOILISVLYKYLEGLTIPSTGPGSIV 238
Db      185 LSGIILQAFGLIRIGFVVYIYSESLISGTTAAHVHLVSQKFTFQLVPSHTDPSIF 244
QY      239 FTFIDICKNLPHTNINSLIFALISGAFVLVYKELNARYMHKIRPPIPTMIYVVAITAIS 298
Db      245 KVLVSFQIEKTNIAIDLVTALIVLVYSIVKEINORFKDKLPVPIPIEFIMTVIAAGVS 304
QY      299 GGCAMPKRYHMOIYGEIQRGFPTPVSPVYSQMKDMIGTAFSLAIVSYINLAMGRTLANK 358
Db      305 YGCDKRNKRYKAVVGDMPGFPPTPVETPQNTVNGDGFIAVMAFAVAFSVASYSLSK 364
QY      359 HGVDVDSNOEMIALGCSNPFSGFEKIHVICALSVTLAVDAGAGSKQVNASLCSLVMT 418
Db      365 YDPLDGNDELIALGLNIVCGVFRGFAGSTALSRSAVOESTGKTQIAGLIGAITIVLIV 424
QY      419 MLVIGIYLYPLKSVYALGALIANLNKSLKQLTDPYLYMRKSLDCCIVWVSFLSSFSL 478
Db      425 VLAIGFLAPLQKSVLALALGNLKMQLMGAIEGRIMRKDYDCLIMTFIIFITVIGL 484
QY      479 PYGAVAGVAFSVLVVYFOTQFRNGYALAOVMDTIYVNPRTYNNRADDIOGKIITTCSP 538
Db      485 GLGLAASVAFOLLITVFTQFPKCSSTLANIGRTNYKKNKDYDMYEPEGVAKIFRCPSPI 544
QY      539 YFANSEIFROKVI-----AKTVSLOELOO-----DPEANP 568
Db      545 YFANIGFRRKLIDAVGSPRLIRLKRKNKALKIRKLOKQGLLOVTPKGFICTVDTIKDS 604
QY      569 PTDPNNO---TPANGTSVSY-ITFSPDSSSPAOSSEBPASAEAPGEPDMLASVPEPV 622
Db      605 DEELDNQIEVLDPQINTDLPFHIDMNDL-----PLNIEVPK-----I 644
QY      623 TPHIILDMSCSYFVLDMLKIKALASLYGKIGVAFVLYNHAOVYNDISHGVEFDGS 682
Db      645 SLSHSLIDFSAVSFLDVSSVRGLKSLIOEFIRIKVYIVGTDDDEFIEKINRYEEF-DOE 703
QY      683 LECKHVPSIHDAVL 697
Db      704 VKSSTIFLTHDAVL 718

RESULT 2
US-08-711-928-2
: Sequence 2, Application US/08711928
: Patent No. 5831015
: GENERAL INFORMATION:
: APPLICANT: SCHEINERST, Clifford W.
: APPLICANT: PAPAS, Takis S.
: TITLE OF INVENTION: Colon Mucosa Gene Having Down-Regulated
: TITLE OF INVENTION: Expression In Colon Adenomas And Adenocarcinomas
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 3000 K Street, N.W., Suite 500
: CITY: Washington, D.C.
: COUNTRY: USA
: ZIP: 20007-5109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25

```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/711,928
: FILING DATE: 11-SEP-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/424,567
: FILING DATE: 17-APR-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/026,045
: FILING DATE: 05-MAR-1993
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 40399/354/NIND
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202)672-5300
: TELEFAX: (202)672-5399
: TELEX: 904136
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 764 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-711-928-2

Query Match      28.9% Score 1120: DB 2: Length 764:
Best Local Similarity 34.0%: Pred No.1.7e-101:
Matches 250: Conservative 142: Mismatches 275: Indels 68: Gaps 11:

QY      7 RYVDRRAVSLTFPDEFEKDRTPVGE-EKLRNARFGSSAKIKAVFGLVLSWLPKY 65
Db      8 QYIVARPVYSTNAFEENHKKTRGHHKTFDLHLKVCSCSPQAKAKIVLSTLFPJASLWPAY 67
QY      66 KIKDYIIPDLGLSGSGSIQVPOGMAFALLANLPAVNGLYSSFFPLTYFFFGVGHQWMP 125
Db      68 RKEWMLSDIYSGISTGIVAVLQGLAFALLVDIPVYGLYASFPAIITLFFGTSRHISV 127
QY      126 GTFVAISILVY----NICQLAPESKFOVE---NNATNESYVDTAAMEARLHVSATLAC 178
Db      128 GFFPLISMVWGLAVGAVSKAVPDRNATTLGLPNNSNSSLDD--ERVRVAAAASYTV 184
QY      179 LTAIOMGLGEMOGFVAIYLSSEFIRGFMTAAGLOILISVLYKYLEGLTIPSTGPGSIV 238
Db      185 LSGIILQAFGLIRIGFVVYIYSESLISGTTAAHVHLVSQKFTFQLVPSHTDPSIF 244
QY      239 FTFIDICKNLPHTNINSLIFALISGAFVLVYKELNARYMHKIRPPIPTMIYVVAITAIS 298
Db      245 KVLVSFQIEKTNIAIDLVTALIVLVYSIVKEINORFKDKLPVPIPIEFIMTVIAAGVS 304
QY      299 GGCAMPKRYHMOIYGEIQRGFPTPVSPVYSQMKDMIGTAFSLAIVSYINLAMGRTLANK 358
Db      305 YGCDKRNKRYKAVVGDMPGFPPTPVETPQNTVNGDGFIAVMAFAVAFSVASYSLSK 364
QY      359 HGVDVDSNOEMIALGCSNPFSGFEKIHVICALSVTLAVDAGAGSKQVNASLCSLVMT 418
Db      365 YDPLDGNDELIALGLNIVCGVFRGFAGSTALSRSAVOESTGKTQIAGLIGAITIVLIV 424
QY      419 MLVIGIYLYPLKSVYALGALIANLNKSLKQLTDPYLYMRKSLDCCIVWVSFLSSFSL 478
Db      425 VLAIGFLAPLQKSVLALALGNLKMQLMGAIEGRIMRKDYDCLIMTFIIFITVIGL 484
QY      479 PYGAVAGVAFSVLVVYFOTQFRNGYALAOVMDTIYVNPRTYNNRADDIOGKIITTCSP 538
Db      485 GLGLAASVAFOLLITVFTQFPKCSSTLANIGRTNYKKNKDYDMYEPEGVAKIFRCPSPI 544
QY      539 YFANSEIFROKVI-----AKTVSLOELOO-----DPEANP 568
Db      545 YFANIGFRRKLIDAVGSPRLIRLKRKNKALKIRKLOKQGLLOVTPKGFICTVDTIKDS 604

```


PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US95/02645
FILING DATE: 01-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0830001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 470 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-471-496-2

Query Match 3.2%; Score 123.5; DB 1; Length 470;
Best Local Similarity 22.3%; Pred. No. 0.0015;
Matches 105; Conservative 70; Mismatches 184; Indels 111; Gaps 21;

QY 35 EKLRAAFRCSSAKIKAVVGLL-----PVL-SWLPRYKIKDYIIPDLGLSG 82
DB 10 QRLKEGRASROLVAVVVALLDNMLFTVVVPVPTFLYMERKE-VISSLHGHS 68
QY 83 SIQVPOGNFALLANLPAVNGLYSFFPLLYFLGCVHOMVGTFAVISILVNICLOL 142
DB 69 S---PHALA-----SPASTIF-SFNNNTV-----AVEESVPSGLAMNDASTI----- 110
QY 143 APESEFOVFNNTNSYVDTAAMEARLHVSATLACLAITOM-----GLGF-- 189
DB 111 -PPPATEAISAKNNCLOGTGFLEETTRVGVLFAS-KAVMOLVNPFGVPLTNRIGYHI 168
QY 190 -MOGEFVAIYLS-----ESFIRGFMTAGLOILSYL----- 220
DB 169 PMFAGFVIMFELSTVWFAFSGYITLLFVARTLOGISSFSVAGLMLASVYDDHERGRA 228
QY 221 --KYIFGLTIPSYGP--GSIVFTFIDICKNLPHTNIASLIFALISGAFLVVKELNARY 276
DB 229 MGTALGGLALGLGVAPGSVMYEVG--KSAPFLILAF--ALLDGAQLOICITQPSKVS 284
QY 277 MHKIR----FPIPTMIYVVAATAISGCKMPKKTTHMOIVGEIORGFPPTVSPVVSQMKD 332
DB 285 PESAGTFLFMLKDPYILVAAGSI---C-----FANMGVALLEPTLPIMMQTCSPKW 336
QY 333 MGTAFSLAIVSYVINLAMGRFLANKHGVDVDSNOEMTALGCSNFGSFFKTHVICALS 392
DB 337 QGLGLFPLPSVSYLIGTNLFGLVANKMGRWLCSLIGMLVGVTS-----LLC 382
QY 393 VTLAVDAG--GKSOVASLCLVLYVMITMLVGLIYLPKSVLGAALAV 440
DB 383 VPLAKHKNGLGIPNAGLGLXIGWESSMMPIMGHVLDPRHNTSVYGSVAHI 432

RESULT 5
US-08-894-840-2
Sequence 2, Application US/08894840
Patent No. 5859200
GENERAL INFORMATION:
APPLICANT: LI, YI
APPLICANT: CAO, LIANG
APPLICANT: ROSEN, CRAIG A.
TITLE OF INVENTION: HUMAN AMINE TRANSPORTER
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005-3934

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,840
FILING DATE: 29-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0830000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 470 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-894-840-2

Query Match 3.2%; Score 123.5; DB 2; Length 470;
Best Local Similarity 22.3%; Pred. No. 0.0015;
Matches 105; Conservative 70; Mismatches 184; Indels 111; Gaps 21;

QY 35 EKLRAAFRCSSAKIKAVVGLL-----PVL-SWLPRYKIKDYIIPDLGLSG 82
DB 10 QRLKEGRASROLVAVVVALLDNMLFTVVVPVPTFLYMERKE-VISSLHGHS 68
QY 83 SIQVPOGNFALLANLPAVNGLYSFFPLLYFLGCVHOMVGTFAVISILVNICLOL 142
DB 69 S---PHALA-----SPASTIF-SFNNNTV-----AVEESVPSGLAMNDASTI----- 110
QY 143 APESEFOVFNNTNSYVDTAAMEARLHVSATLACLAITOM-----GLGF-- 189
DB 111 -PPPATEAISAKNNCLOGTGFLEETTRVGVLFAS-KAVMOLVNPFGVPLTNRIGYHI 168
QY 190 -MOGEFVAIYLS-----ESFIRGFMTAGLOILSYL----- 220
DB 169 PMFAGFVIMFELSTVWFAFSGYITLLFVARTLOGISSFSVAGLMLASVYDDHERGRA 228
QY 221 --KYIFGLTIPSYGP--GSIVFTFIDICKNLPHTNIASLIFALISGAFLVVKELNARY 276
DB 229 MGTALGGLALGLGVAPGSVMYEVG--KSAPFLILAF--ALLDGAQLOICITQPSKVS 284
QY 277 MHKIR----FPIPTMIYVVAATAISGCKMPKKTTHMOIVGEIORGFPPTVSPVVSQMKD 332
DB 285 PESAGTFLFMLKDPYILVAAGSI---C-----FANMGVALLEPTLPIMMQTCSPKW 336
QY 333 MGTAFSLAIVSYVINLAMGRFLANKHGVDVDSNOEMTALGCSNFGSFFKTHVICALS 392
DB 337 QGLGLFPLPSVSYLIGTNLFGLVANKMGRWLCSLIGMLVGVTS-----LLC 382
QY 393 VTLAVDAG--GKSOVASLCLVLYVMITMLVGLIYLPKSVLGAALAV 440
DB 383 VPLAKHKNGLGIPNAGLGLXIGWESSMMPIMGHVLDPRHNTSVYGSVAHI 432

RESULT 6
US-09-139-675-2
Sequence 2, Application US/09139675A
Patent No. 6117426
GENERAL INFORMATION:
APPLICANT: LI, YI
APPLICANT: CAO, LIANG
APPLICANT: ROSEN, CRAIG
TITLE OF INVENTION: Human Amine Transporter
FILE REFERENCE: 1488.0830003
CURRENT APPLICATION NUMBER: US/09/139,675A
CURRENT FILING DATE: 1998-08-25

EARLIER APPLICATION NUMBER: WO PCT/US95/02645
 EARLIER FILING DATE: 1995-03-01
 EARLIER APPLICATION NUMBER: US 08/471,496
 EARLIER FILING DATE: 1995-06-06
 NUMBER OF SEQ ID NOS: 9
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 2
 LENGTH: 470
 TYPE: prt
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: UNSURE
 LOCATION: (402)
 OTHER INFORMATION: May be any amino acid
 US-09-139-675-2

Query Match	3.2%	Score 123.5	DB 3	Length 470
Best Local Similarity	22.3%	Pred. No. 0.0015		
Matches 105	Conservative 70	Mismatches 184	Indels 111	Gaps 21

```

QY      35 EKLNNARFCSASAKKAVFGL-----PVL-SMLPKYIKDYIIPDLGLGSLG 82
      10 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      10 QRLKEGASQQLVYVYFVALLDNNLFTVYVPIVTFLLDMFEKE-VTSSLHGAGS 68

QY      83 STQYQCMAPALLANLPAVNGVYSFPPLTYFFLGCVHOMPCETFAVISILVGNICLQ 142
      69 S---PHALA-----SPAFTIF-SFENNMTV-----AVEESVPGIAMMDTASTI--- 110
Db

QY      143 APESKFOVENNATNESYVDTPAAMEARLHASATLACLTAIHOM-----GLGF-- 189
Db      111 -PPATEAISHAKNNCLQGTGFEETETRVGVLEAS-KAAMQLLVNPFGPLNRIGTHI 168

QY      190 -MOEGFAIYLS-----ESPIRGFMTAGLOILISVL----- 220
      169 PMFAGFVIMFLSTWMPAFSGTYTLFPARTLQIGGSSFSSVAGIGMLASYTDDHENGRA 228
Db

QY      221 --KRIEGLTIPSTYGP--GSIVTFEFDIDCKNLPHTNINASLIFALISAFVLVKELNARY 276
      229 MGTALGALGALGLVGAEPGSVMYEFVG--KSAPLLIATFL--ALLDADLOCLIOPSKVS 284
Db

QY      277 MHKIR-----PIPTPEMYVVVAATISGCKMKPKYHQIYGEIDGRPPTVPSVPSQWD 332
      285 PESAKGTPFLMLLDPIYIIVAAGSI---C-----FAMKGAILEPTLPIMMQTMCSPKW 336
Db

QY      333 MGTAFSLAIVSYVINLAMEFTLANKHGYDVDSNOEMIALGCSNFSGSPFKIHVICALS 392
      337 QLGIALEFPAISVSYLTIGNLFCVLANKKGRMCLSLIGMLVVGTS-----LLC 382
Db

QY      393 VTLAVDAG--GRSQVASLCSLSLVMTMLVLGTYLPPLKRSVGLALIAV 440
      383 VPLANKFGLGPNAGLGLXIGWESSMMPITGMLVDPDRHTSYVGSVAHI 432
Db

```

```

1 RESULT 7
2 US-09-110-116-3
3 : Sequence 3, Application US/09110116
4 : Patent No. 6013479
5 : GENERAL INFORMATION:
6 : APPLICANT: Xu, Hong
7 : APPLICANT: Cohan, Victoria L.
8 : APPLICANT: Stuart, Susan G.
9 : TITLE OF INVENTION: HUMAN EMBL-LIKE G PROTEIN COUPLED
10 : TITLE OF INVENTION: RECEPTOR
11 : FILE REFERENCE: PR-0550 US
12 : CURRENT APPLICATION NUMBER: US/09/110,116
13 : CURRENT FILING DATE: 1998-07-02
14 : NUMBER OF SEQ ID NOS: 4
15 : SOFTWARE: FastSeq for Windows Version 3.0
16 : SEQ ID NO 3
17 : LENGTH: 886
18 : TYPE: PRT
19 : ORGANISM: HOMO SAPIENS

```

; FEATURE:
 ; OTHER INFORMATION: 784994, GenBank
 US-09-110-116-3

Query Match	2.7%	Score 104.5	DB 3	Length 886
Best Local Similarity	20.7%	Pred. No. 0.32		
Matches 63; Conservative	51;	Mismatches 105;	Indels 85;	Gaps 14.

```

OY 264 AFVLVLVELNARUYNHKIRPPIRPEMIIVVAVATISGGCKMKRYKMYOIVEIOLQGFPTPV 3323
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 482 SFVGSVSLNERFQDDHAPLTTTSEIKLKNRSVGGI-----MTGCKDGFSDPI 5522
OY 324 -----SPVVSQMK-DMIG--TAFSLAIV--SYVI-----NLAM-- 3515
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 533 IYTLNNQPKOKFERPLCYSMTSDYVKGGRMTSPBCVILLESSEYTTTISCNMNLAVIYA 5929
OY 352 -GRTLANKHGYDVDSNOEMIALGC-----SNFF-----GSFEKTIHV-ICCALSVTLA 3965
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 593 SGELTMDPFSLYIIISHVGIISLVCGLVALAIFELLCRSIRNNHNTYLHLCLVCLLAKTLE 6524
OY 397 VDGAGGGSQVAAISLCSVLVMTMLVLCIYLPLPKSVLGGALIAVNLKSLKOLTDPYYLM 4565
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 653 LAGI-HKTDNKTGCAGIILAGFLHYELFCFMMLEAVEAIIELFLMVRNLK----- 6989
OY 457 RKSLDCCIWMVSPFLSEFPLSLPGVAVGVAFSVLVVPQOQR-NCYALAQVMDTDIYV 5119
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 699 -----VNVFSSRNIMKLHICAFGIGLPLMVVVYISASVQPGSIG-----MHNRCWL 7444
OY 516 NPKT 519
      : : :
Db 745 NTEI 748

```

RESULT 8
US-08-843-417-10
Sequence 10, Application US/08843417
Patent No. 6184349
GENERAL INFORMATION:
APPLICANT: Herman, Ronald C
APPLICANT: Delgado, Stephen G
APPLICANT: Fish, Linda M
APPLICANT: Sangameswaran, Lakshmi
APPLICANT: Rabert, Douglas K
TITLE OF INVENTION: CLONED PERIPHERAL NERVE
TITLE OF INVENTION: TETRODOXIN-RESISTANT SODIUM CHANNEL alpha-SUBUNIT
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Heller Ehrman White & McCauliffe
STREET: 525 University Ave
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.A.
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/843,417
FILING DATE: April 15, 1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Schmonsees, William
REGISTRATION NUMBER: 31,796
REFERENCE/DOCKET NUMBER: 28340-P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)-324-7041
TELEFAX: (415)-324-0638
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1956 amino acids

TYPE: amino acid
 STRANDEDNESS: unknown
 TOPOLOGY: unknown
 MOLECULE TYPE: peptide
 HYPOHETICAL: NO
 ANTI-SENSE: NO
 US-08-843-417-10

Query Match 2.7%; Score 103.5; DB 4; Length 1956;
 Best Local Similarity 19.4%; Pred. No. 1.5;
 Matches 112; Conservative 93; Mismatches 244; Indels 127; Gaps 22;

QY 122 QMVGTEFVITLVGNICQLAPESKFOVENNATNESYVDTAAMEARLHSATLA--CL 179
 DB 576 ELAPGAVDVSAFDAGCKTFLSAEFLDEPFRAGRAMSVSITTSVLEELSECKCPCL 635
 QY 180 TAIIO-----MGLGFQRFVAIYLSSESTIRGFMTAGIQLISVLYKF---- 224
 DB 636 TSLSKYLIWDCPCPMVKLTFLGLVTDPEFELTI-----TLCIVNTTFMAME 685
 QY 225 --GL--TIPSYTGPSSIVFTFIDICKNLPHTNINSLIFALISGAFVLVKELNARYMKI 280
 DB 686 HHGMSPTFEAMIQIGNIYFTIF-----FTAEVFKIITA-----*FDPIYYFOK 727
 QY 281 RPIPTFEMIIVV---VATPISGCKMPKKYHMOIVGEIQGFPPTPVSPVVSQMKDMIGT 336
 DB 728 KWNIFDCIIVTVSLEELGVAKKGSLSVLRSPFLRVFKLAKSMPT-LMTLKIKIGNSVGA 786
 QY 337 ---AFSLAIYSVINLGMRTLANKHGYDVDSNOEMIAL-----GCSNFPSGFKRI 384
 DB 787 LGNLTIIILITIVFVAL--VGKQL--GENYRNNRKNISAPHEPMRHHMHDFHSFLIV 842
 QY 385 HVICCALSVTLAVDAGGSKQVASLCSLVVMTLVIGIYLYPLKSVLGLALIVNLKN 444
 DB 843 FRILCG---EMENMMAMOEYCKSICLILEFTYVWGNLV-----VLNFIALLNS 892
 QY 445 -SLKQITDPTIYLMKRSKIDCCIVVSF-----LSSFF-LSLPYGAVGAVSFLVVV 494
 DB 893 FSADMLTAPEDDEGVNMLQVALRIQVGHRTKQALCSFFSRSCFPPOPKAPELVLVLP 952
 QY 495 FOTFRNGYALQVM-----DIDIYVNPRTYNNRQDLOGIKIITYYCSPL 538
 DB 953 LSSSAENHIANTRARGSSGGLQAPRGPRDEHSDFIANPTVAVSPIAGESEDLDLEDD 1012
 QY 539 YFANSEIFRQKVIAT--TVSLQELQODENAPPTDPNNQTPANGTSVYITFSPDSSSP 596
 DB 1013 GGEPAOSFOQEVYIPKGOEQLOQVRECGDHLTP-----RSPGTGTSSDDLAPSLGETWK 1066
 QY 597 AQSEPPASAPG-----EPSDMIASVP 619
 DB 1067 DESVQAPAEGVDDTSSSEGSTVDCLDPEILRKIP 1102

RESULT 9
 US-08-677-049-9

Sequence 9, Application US/08677049
 Patent No. 5858707
 GENERAL INFORMATION:
 APPLICANT: Guimaraes, M. Jorge
 APPLICANT: Bazan, J. Fernando
 APPLICANT: McCleanahan, Terrell K.
 TITLE OF INVENTION: PURIFIED MAMMALIAN NUCLEOBASE PERMEASES;
 TITLE OF INVENTION: NUCLEIC ACIDS; ANTIBODIES
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: DNA Research Institute
 STREET: 901 California Avenue
 CITY: Palo Alto
 STATE: California
 COUNTRY: USA
 ZIP: 94304-1104

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/677,049
 FILING DATE: 03-JUL-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/000,788
 FILING DATE: 03-JUL-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Ching, Edwin P.
 REGISTRATION NUMBER: 34,090
 REFERENCE/DOCKET NUMBER: DX0511
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-852-9196
 TELEFAX: 415-496-1200
 INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 438 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 FEATURE:
 NAME/KEY: Region
 LOCATION: 115..144
 OTHER INFORMATION: /note= "Encompasses TM 4 of Figure 4"
 FEATURE:
 NAME/KEY: Region
 LOCATION: 323..357
 OTHER INFORMATION: /note= "Encompasses TM 9 of Figure 4"
 FEATURE:
 NAME/KEY: Region
 LOCATION: 359..386
 OTHER INFORMATION: /note= "Encompasses TM 10 of Figure 4"
 OTHER INFORMATION: 4"
 US-08-677-049-9

Query Match 2.7%; Score 103; DB 2; Length 438;
 Best Local Similarity 19.6%; Pred. No. 0.14;
 Matches 80; Conservative 76; Mismatches 132; Indels 120; Gaps 22;

QY 172 VSATLACITAIOMKLGMOGFVAIYLSSESTIRGFMTAGIQLISVLYKFGLTIPSY 231
 DB 74 LGCTFTAVSPMAIG--SEYGVSTV-----GSIIASG--ILVILISFEFKLV-SF 120
 QY 232 TGP---GSIV---FTFIDICKNLPHTNIA-----SLIFALISGAFL----- 266
 DB 121 PFPVVTGSAVVTITGITLMPVAMN---NMAGEGADGDSLNLATFTVSIILYRF 176
 QY 267 -----VLKELNARYMKIRPPIPTFEMIIVVAVATPISGCKMPKKYHMOIVG 313
 DB 177 TKGFIKSVSILIGLIGFIAYFMCKVOFDNVSDAAV-----QMIO 218
 QY 314 ELQKGFT-PPSPVVSQMKDMIGTAFSLAIVSY-----INLGMRTLANKHGYDVDSNOE 368
 DB 219 PFFGAPSFHAAPIT-----TMSIYAIVSLVESTGCVFALG-DLTNRRLTEIDLSKG 269
 QY 369 MIALGCSNFGSPFKIHVICCALSVTLAVDAGGSKQVASLCSLVVMTLVIGIY-LY 427
 DB 270 YRAGSLAVLGGTIN-----AFPTT-AFSQNVGLVQITGKKNAYIVTVYIIMAGELF 322
 QY 428 P-----LPKSYLG---ALIVANLKNLSKQITDPTIYLMRKSRL--DCCIVV---VSF 470
 DB 323 PKIAFTTIIPISAVLGGAMVAMFGVIAVGIKMLSRIDFAQENMLLIYACSGVGLGVTY 382
 QY 471 LSSFFLSLP-----YGAVGAVSFLVVVFOFTRNGYALQVMD 510

ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-6408A2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-396-001-4

Query Match 2.7% Score 103; DB 2; Length 834;
Best Local Similarity 17.6% Pred No. 0.41; Indels 250; Gaps 33;
Matches 140; Conservative 110; Mismatches 294;

QY 93 ALLANLPAVNGLYSFFPLTYFPL--GGVHQMVPQTFAVI-----SILVGNICL 140
DB 87 ASMAMPAL-----LPLINEFDLEMDGPRKSSHDFTVAPNSGVNTSSLIMETPS 139
QY 141 QLAPEKQOVF---NNATNESYVDTAAEAERLHVSATLACTAITIOMGLGM-----OF 192
DB 140 SVTPAASLRNFNSNNAAKSCGVNDSFGLSS--STSSSWEISALPLRLDYIKLATDOF 198
QY 193 G--FVAIYL---SES-FIRGEFTAGLOLISLVKYGTLTIPSTGPGSIYFTFIDCK 246
DB 199 GCRFLQKLETPESESNNVADLMYEQIKPFLLILDPEF-----NLYOKLCD 246
QY 247 NLPHNTIASLIFALISGAFVLVYKELNARYMKIRFPITPTMIVVAVATAISGCKMPKK 306
DB 247 YLTAEQKTLTIOTIYPNVFOISINOYGRSLQKIIDTVONEQDILLI----- 294
QY 307 YHMQIVGELQKGF--TPVSPVVSQMKDMGTAFSLAIVSYVILNMGRTLANKHGIDV 363
DB 295 -----KGFSEFTSIEQVTVLINDLNG-----NHVIOKCIJFKFSPSKFGFTI 336
QY 364 DS--NOEMIALGCSNPFSGFFKIHVICALSVTLAVDAGAGSOVASLCVSLVVMITMLV 421
DB 337 DAIVDQNNITITSTHKHG-----CCVLOKLLSVCTL---QOIFKISVKIVQFLPGLI 385
QY 422 ---LGIIVLYPLPKSV-----LGAIIAIVLNKLSKQLDPPYLYMKRSKLDCCIMVVSFLS 472
DB 386 NDQFNNTIIOFLDKELDEYLLAEFLNRLSNELCOLS-----CLKRSSNVV 432
QY 473 SFFLSLPYGVAVG-----ATSV-LVVVFQTOF 499
DB 433 EKFIKKLRITITGTVNNNGASQRTAAVADVDVINAAMNILLTTIDFTVNLNVLIRDF 492
QY 500 RRGVYLAQVMDTIYVNPRTYNNRAODIOG----- 528
DB 493 GN-YALQTLLDVKNYSPLLAIVNKNNSNAIGONSSSTLNGFNCNDSLSKIGNILVLTKELL 551
QY 529 -----IKTIYCSPLYFANSEIFROKVIYAKTVSLQELQODFEN----- 566
DB 552 PSIKTTYSARKIKELVKVAAEATGIPFTDISPOVTAMSHNNQITINNEKNKNHNSHNH 611
QY 567 -----APPTDPYNNQ-----TPANG---TSVSYIT--ESPSSSPAQSEPPA-- 603
DB 612 NHHNNHNNHNNNNNNNNKSHTRHPSLPAANAHRSSNSSVTNFNSNOYAADOKIKHSPDI 671
QY 604 -----SADAPGEPSPDMLASVPVTFPHLLIDMSGVSPVDLMGKALAKLSSTYKIGIV 657
DB 672 MNEFONAIYPSMGAPSFNSQTNPLVYS--HNSLQNFQNRQANLMAHPNSAAPLHSSSSNI 730
QY 658 KVLVNIHQAQYNDISHGCVF--EDGSLECKHVFPSIDAVLFAQANARDVTPG----- 709
DB 731 ---TNVAPNYSRGFGKQGFMMNETDKINAHNFS-----YSNANQONFNESEVPRMQ 779
QY 710 HNFQAGAPDAELSL 723

DB 780 YQTEGANMDSLSLM 793

RESULT 12
US-09-323-433A-4
Sequence 4, Application US/09323433A
Patent No. 6218512
GENERAL INFORMATION:
APPLICANT: Guarante, Leonard P.
APPLICANT: Austriaco Jr., Nicanor
APPLICANT: Claus, James J.
APPLICANT: Cole, Francesca
APPLICANT: Kennedy, Brian
TITLE OF INVENTION: GENES DETERMINING CELLULAR SENESENCE IN
FILE REFERENCE: 0050.1491-003
CURRENT APPLICATION NUMBER: US/09/323,433A
CURRENT FILING DATE: 1999-06-01
PRIOR APPLICATION NUMBER: US 08/396,001
PRIOR FILING DATE: 1995-02-28
PRIOR APPLICATION NUMBER: PCT/US94/09351
PRIOR FILING DATE: 1994-08-15
PRIOR APPLICATION NUMBER: US 08/107,408
NUMBER OF SEQ ID NOS: 48
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 834
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-323-433A-4

Query Match 2.7% Score 103; DB 4; Length 834;
Best Local Similarity 17.6% Pred No. 0.41; Indels 250; Gaps 33;
Matches 140; Conservative 110; Mismatches 294;

QY 93 ALLANLPAVNGLYSFFPLTYFPL--GGVHQMVPQTFAVI-----SILVGNICL 140
DB 87 ASMAMPAL-----LPLINEFDLEMDGPRKSSHDFTVAPNSGVNTSSLIMETPS 139
QY 141 QLAPEKQOVF---NNATNESYVDTAAEAERLHVSATLACTAITIOMGLGM-----OF 192
DB 140 SVTPAASLRNFNSNNAAKSCGVNDSFGLSS--STSSSWEISALPLRLDYIKLATDOF 198
QY 193 G--FVAIYL---SES-FIRGEFTAGLOLISLVKYGTLTIPSTGPGSIYFTFIDCK 246
DB 199 GCRFLQKLETPESESNNVADLMYEQIKPFLLILDPEF-----NLYOKLCD 246
QY 247 NLPHNTIASLIFALISGAFVLVYKELNARYMKIRFPITPTMIVVAVATAISGCKMPKK 306
DB 247 YLTAEQKTLTIOTIYPNVFOISINOYGRSLQKIIDTVONEQDILLI----- 294
QY 307 YHMQIVGELQKGF--TPVSPVVSQMKDMGTAFSLAIVSYVILNMGRTLANKHGIDV 363
DB 295 -----KGFSEFTSIEQVTVLINDLNG-----NHVIOKCIJFKFSPSKFGFTI 336
QY 364 DS--NOEMIALGCSNPFSGFFKIHVICALSVTLAVDAGAGSOVASLCVSLVVMITMLV 421
DB 337 DAIVDQNNITITSTHKHG-----CCVLOKLLSVCTL---QOIFKISVKIVQFLPGLI 385
QY 422 ---LGIIVLYPLPKSV-----LGAIIAIVLNKLSKQLDPPYLYMKRSKLDCCIMVVSFLS 472
DB 386 NDQFNNTIIOFLDKELDEYLLAEFLNRLSNELCOLS-----CLKRSSNVV 432
QY 473 SFFLSLPYGVAVG-----ATSV-LVVVFQTOF 499
DB 433 EKFIKKLRITITGTVNNNGASQRTAAVADVDVINAAMNILLTTIDFTVNLNVLIRDF 492
QY 500 RRGVYLAQVMDTIYVNPRTYNNRAODIOG----- 528
DB 493 GN-YALQTLLDVKNYSPLLAIVNKNNSNAIGONSSSTLNGFNCNDSLSKIGNILVLTKELL 551

```

OY 529 -----IKITYCSPLYEANSIEIFROKVIYAKTVSLOELOODEFEN----- 566
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 552 PSIKTTSYAKIKIKYKAYAEATGIPFTDISPOVTAMSHNNIQTINNEKNBNHKNSHNH 611
OY 567 -----APPTDPNNNO-----TPANG---TSVSYIT---FSPDSSSPQOSEPPA--- 603
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 612 NNNHHNNHHNNNNNNNNQSSHTRHESLPANAYHRRSSSVTNNSQVADQKIHSPQOI 671
OY 604 -----SAPGEPSDMLASVPEVFEFTLLIDMSGVSFVDLMGIALAKLSTYGIKV 657
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 672 MNFNQNAVYSKMGAPSPNSQTNPLVS-HNSLQNFDRKOPANLMAHNSAAPHSESSNI 730
OY 658 KVELVNIHAQVYNDISHGVF---EDGSLCKHVPSPSIDAVLFAQANARDVTPG----- 709
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 731 ---TNVNNVNSRGFEKQPGFMNNETDKINAHNSP-----YSNANSQNFNESFVPRMQ 779
OY 710 HNFQCAPGDAELSL 723
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 780 YQTEGANMDSLSM 793

```

```

RESULT 13
US-08-132-990A-4
; Sequence 4, Application US/08132990A
; Patent No. 5834589
; GENERAL INFORMATION:
; APPLICANT: MERUELO, DANIEL
; APPLICANT: YOSHIMOTO, TAKAYUKI
; TITLE OF INVENTION: Human Retrovirus Receptor and DNA Coding Therefor
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESS: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/132,990A
; FILING DATE: 07-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/084,729
; FILING DATE: 29-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/05569
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/899,075
; FILING DATE: 11-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/806,178
; FILING DATE: 13-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/627,950
; FILING DATE: 14-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 8105-004-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66441 PENNIE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 622 amino acids
; TYPE: amino acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-132-990A-4

```

```

Query Match 2.6%; Score 101; DB 2; Length 622;
Best Local Similarity 20.5%; Pred. No. 0.4;
Matches 96; Conservative 60; Mismatches 206; Indels 106; Gaps 17;

```

```

OY 97 NLPVAVGLVSSPEPLLTFFELGIV-----HQMVGEFAVISILV-----GNI 138
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 155 NAFVLAQTPDIPFVAVIIIIITGLTLGVKESAMVKKFTCNVAVLCIVVSGEYKGS 214
OY 139 CLOLAPESKFOVENNATNESYVDFAAMEARLHV-SATLACLTAIIOMIGFMQGFVAI 197
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 215 KMNQLEKNKFNPCNNNDTNVRYGEGFMPPGFSGVLSGANTCFYAFV-----GFD 265
OY 198 YLSESTIRGEMTAAGLIQLISVLKYIFGLTIPSTPGSIV-----FTFIDCKNP--- 249
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 266 ATGGEVKNPKAKAPVGIVASL-----ICFTAYFGVSALTLMMPYFCIDIDSPLP 320
OY 250 -HTNIASLIFALISGAFIVLVKELNARYMHKIFPPIPTETIVVVAFTAISG-----GCM 303
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 321 KHQWBEAKYAAVAGSLCALSTSLGS-----MFPMPR-----YIYMAEDGLFKFLAKI 371
OY 304 PKRYHMOIVGEIORGFPPTSPVVSQKDM-----IGTAFSLAIVSYI-----N 348
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 372 NNRTKTPVATVYSAGIAAVMAFLFELKDLVDMSIGTLLAVSVACVLYLVKQPEGRN 431
OY 349 LAMGRILANKHGYDVDSNOEMIA-----LGCNFFGSGFFKI 384
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 432 LYYQMARTEELDRVDONELVNASSESOTGFLPYAEKFSLSKLSLSPKNVPSKSGILVINI 491
OY 385 HV-ICCALSVTLAVDAGGSGQVASLCSLV-----VMITMLVLGIYLPPLPKVLGAL 437
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 492 SAGLLAALIIITVCIYAVLGRALABEGLTNAVFMTSGSVLLCMVLTGI-IMROPESTKIS 550
OY 438 IAVNLKNSLKQLTPPYLWRKSKIDCCIYV--VSFLSFFLSLPYGV 482
Db 551 FKVFVFPVLPVLSIFVNIYLMQDLDOGTWVRFVMMVLIGFTIYFGGI 598

```

```

RESULT 14
PCT-US92-09382-4
; Sequence 4, Application PC/TUS9209382
; GENERAL INFORMATION:
; APPLICANT: MERUELO, DANIEL
; APPLICANT: YOSHIMOTO, TAKAYUKI
; TITLE OF INVENTION: Human Retrovirus Receptor and DNA Coding
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Broadway and Nelmark
; STREET: 419 Seventh Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/09382
; FILING DATE: 19921213
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Livanat, Shmuel
; REGISTRATION NUMBER: 33,949
; REFERENCE/DOCKET NUMBER: MERUELO-1
; TELECOMMUNICATION INFORMATION:

```

TELEPHONE: 202 628-5197
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 622 amino acids
 TYPE: AMINO ACID
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US92-09382-4

Query Match 2.6%; Score 101; DB 5; Length 622;
 Best Local Similarity 20.5%; Pred. No. 0.4;
 Matches 96; Conservative 60; Mismatches 206; Indels 106; Gaps 17;

QY 97 NLPVNGLSSEFFLLTYFLGVS-----HQMVGCTFAVISILY-----GNI 138
 DB 155 NMPGVLQPDIFAVIIIIITGLTLGVKESAMVKIFTCINVLVLCFLVSGFVKGI 214
 QY 139 CLQLAPESKFOVFNNATNESYVDTAAMEARLHV-SATLACLTAITOMGLGFMOGFVAI 197
 DB 215 KNNQLEKMFSCNNNDNTNKKYGGGFMPPFGFSGVLSGATCFYAFV-----GFDGI 265
 QY 198 YLSESEFIRGFMPAAGLIILISVLKYIFGLTIPSYTGPISIV-----FTFIDICKNP--- 249
 DB 266 ATGGEVKNPKOKAIPVGIASLL-----ICFIYFGVSALTLMPYFCIDIDISPLPGAF 320
 QY 250 -HTNINSLIFALISGAFVLVVELNARMKIRFPIPTKIRIVVVAFTASG-----GCKM 303
 DB 321 KIQGWEAKYVAAGISGLALSTSLGS-----MFPMPR-----VIYMAEDGLLKFELAKI 371
 QY 304 PKYHQIYGEIQRGFPTVPVSPVSOVKDM-----IGTAFSLAIVSYI-----N 348
 DB 372 NNRTKTPVATATYSGAIANVMAFLFELKDLVDLMSIGTLLASLVAACVLVRQEPQR 431
 QY 349 LAMGRTLANKHGYDVDSNOEMIA-----LCCSNFEGSGFFKI 384
 DB 432 LVMQARTTEELDRVDONELVSSASESQTGFLPVAEKFSLSKLSIPKRVSPSKFGLIVNI 491
 QY 385 HV-ICGALSVTLAVDAGSKSOVASLCSLV-----VMITMVLGILYLYPLPKSVLGL 437
 DB 492 SAGLALALITVYICVAVLRELAEGTLMAVFMVMTGSLVLCMLVTI--IMROESKTKLS 550
 QY 438 IAVNLKSLKQLTDPYLMRKSKLDCIIV--VSFLSFFLSLPGV 482
 DB 551 FVVPFVPLPVLSIFVNIYLMQLDGTWVRFAVMMLIGFTIYFGIGI 598

RESULT 15
 US-08-898-976-2
 Sequence 2, Application US/08898976
 Patent No. 5891670

GENERAL INFORMATION:
 APPLICANT: Burnham, Martin
 APPLICANT: Lonetto, Michael
 APPLICANT: Warren, Patrick
 TITLE OF INVENTION: NOVEL TETRACYCLINE RESISTANCE
 TITLE OF INVENTION: PROTEIN
 NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Dechert Price & Rhoads
 STREET: 997 Lenox Drive, Building 3, Suite 210
 CITY: Lawrenceville
 STATE: NJ

COUNTRY: USA
 ZIP: 08543

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/898,976

FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Bloom, Allen
 REGISTRATION NUMBER: 29,135
 REFERENCE/DOCKET NUMBER: GM10044
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 609-520-3214
 TELEFAX: 609-520-3259
 TELEX:

INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 462 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear

US-08-898-976-2

Query Match 2.6%; Score 99.5; DB 2; Length 462;
 Best Local Similarity 17.7%; Pred. No. 0.34;
 Matches 83; Conservative 72; Mismatches 121; Indels 193; Gaps 19;

QY 110 PLTYFELGIVHMQVGTFAVISILVGNICQLAPESKFOVFNNATNESYVDTAAMEAR 169
 DB 25 PILLELFVPSL-VIDNSKRLISVAI-----ADD 51
 QY 170 LHVASATL-----ACLTAAIOMGLGFMOGFV--AIYLSSEFIRGF----- 207
 DB 52 LINSVTVSMQATLAV-IGMGAVVYASLSDAISIRPPFIYGVLLIFGSIIGNIFQH 110
 QY 208 -----MTAAGLIILISVLKYI-----FGLTIPST-----TGCGSIV 238
 DB 111 SPLTLVGRITOTAGLAAETLVYIVYAKYLSKEDOKTYGLSTSSYSLSLVIGTISGFI 170
 QY 239 FPFID-----ICKNLP--HTNINSLIF-----ALISGAFVLVREL 272
 DB 171 FTYLHMTNMFLLALIVFTLPFLFKLPKKNNTKKAHLDFVGLLVAITATVMTLFTNF 230
 QY 273 N-----ARYMKIRFPIPT-----MIVVVAATASGG-----C 301
 DB 231 NMLYMGALITATVYIKNAQRPVKNKSEFQNKRYASFLLFVGVAAIQIGYFTFPE 290
 QY 302 KMPKRYHMQI-----VGEIQGFPTPPVSPVSOVKDMIGTAF 338
 DB 291 IMEQIYHQLDPTSLLPVGYIVAVIGALSGKIGEYLNKQAIITAILLALSLILPAF 350
 QY 339 SLA-IVSYVINL-----AMGRTLANKHGYDVDSNOEMIALCSENFSGSPFKIHVC 388
 DB 351 AVGNHISIFVISMIFPAGSALMAYAPLLEAKITIDLNMGTGAIGYVNL-----IIN 402
 QY 389 CALSVTLAVDGA-----GKNSOASLCSLVMTMLVLGILY 427
 DB 403 VAVSVGIAIAAALIDKALNFGNDALSSHFGIILITLMSIVGLVF 451

Search completed: April 26, 2002, 09:05:37
 Job time: 203 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 09:01:49 ; Search time 28.07 Seconds
(without alignments)
1987.075 Million cell updates/sec

Title: US-09-749-589-2
Perfect score: 3869
Sequence: 1 MSQPRRYVVDRAAYSLTLF.....MDLEQEMFGSMFHAETLAL 753

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /SIDSB8/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SIDSB8/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SIDSB8/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SIDSB8/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SIDSB8/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SIDSB8/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SIDSB8/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SIDSB8/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SIDSB8/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SIDSB8/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SIDSB8/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SIDSB8/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SIDSB8/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SIDSB8/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SIDSB8/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SIDSB8/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SIDSB8/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SIDSB8/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SIDSB8/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SIDSB8/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SIDSB8/gcgdata/geneseq/geneseq/AA2000.DAT:*
22: /SIDSB8/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1259.5	32.6	758	21	AAV71067 Human membrane tra
2	1116	28.8	739	15	AA860568 Down-regulated in
3	974.5	25.2	739	20	AAV08928 Rat DTDSF protein.
4	972	25.1	739	20	AAV08928 Human DTDSF protei
5	735.5	19.0	506	22	AAAM5272 Peptide #1706 enco
6	735.5	19.0	506	22	AAAM2737 Peptide #1774 enco
7	735.5	19.0	506	22	AAAM3017 Peptide #1699 enco
8	601.5	15.5	143	22	AAAM2394 Human polypeptide
9	576	14.9	593	21	AAV44945 wheat sulphate per
10	566	14.6	680	21	AAV44943 Soybean sulphate p
11	566	14.6	688	21	AAV44935 Corn sulphate perm

12	549	14.2	631	21	AA629247 Arabidopsis thalia
13	543.5	14.0	658	21	AA631621 Arabidopsis thalia
14	543.5	14.0	658	21	AA632200 Arabidopsis thalia
15	543.5	14.0	703	21	AA631620 Arabidopsis thalia
16	538.5	13.9	621	21	AAV44942 Soybean sulphate p
17	537	13.9	656	21	AAV44944 Wheat sulphate per
18	498	12.9	579	21	AAV44936 Corn sulphate perm
19	452.5	11.7	842	21	AAV44358 P. chrysogenum sut
20	437	11.3	746	21	AAV44358 Corynebacterium su
21	425	11.0	565	22	AAAB76817 C glutamicum prote
22	425	11.0	579	22	AAAG90917 Human transporter
23	403.5	10.4	606	22	AAE04904 Arabidopsis thalia
24	385.5	10.0	499	21	AAAG29248 Arabidopsis thalia
25	355.5	9.2	510	21	AAAG31622 Arabidopsis thalia
26	355.5	9.2	510	21	AAAG32201 Arabidopsis thalia
27	332.5	8.6	446	21	AAAG29249 Arabidopsis thalia
28	288	7.4	457	21	AAAG32202 Arabidopsis thalia
29	285	6.6	483	22	AAAG3019 S. epidermidis ope
30	253	6.5	537	22	AAAG90717 Corynebacterium g
31	253	6.5	537	22	AAAB76818 Human secreted pro
32	240	6.2	76	21	AAAG03678 Human OREF ORF84 p
33	237.5	6.1	128	21	AAAB40320 Cosmid CHR15 enco
34	225	5.8	466	21	AAV95728 Corn sulphate perm
35	220.5	5.7	311	21	AAV44937 H. pylori GPO 141
36	201	5.2	390	19	AAAG8528 S. epidermidis ope
37	191	4.9	411	22	AAAG2493 C glutamicum prote
38	179.5	4.6	485	22	AAAB76816 Corynebacterium g
39	179.5	4.6	485	22	AAAB7557 Human OREF ORF951
40	172.5	4.5	312	21	AAAB07557 Protein encoded by
41	170	4.4	582	21	AAAB07557 Peptide #6232 enco
42	163	4.2	55	22	AAAM19798 Peptide #7257 enco
43	163	4.2	55	22	AAAM33220 Chlamydia trachoma
44	161	4.2	416	20	AAV37263 Artichoke sulphate
45	151.5	3.9	226	21	AAV44938

ALIGNMENTS

RESULT 1	AAV71067	standard; Protein: 758 AA.
ID	AAV71067	
XX	AAV71067	
AC	AAV71067	
XX	AAV71067	
DT	29-AUG-2000	(first entry)
XX	AAV71067	
DE	Human membrane transport protein, MTRP-12.	
XX	Human membrane transport protein, MTRP-12; antiinflammatory; cytosolic; antithyroid; immunosuppressive; thymimetic; antidiabetic; nootropic; antiarthritic; neuroprotective; antidepressant; nephrotoxic; virucide; antihelminthic; protozoacide; antibacterial; neuroleptic; antiout;	
KW	diagnosis; prevention; treatment; membrane transport disorder; epilepsy;	
KW	Alzheimer's disease; diabetes; Parkinson's disease; neurological disorder;	
KW	inflammatory disorder; AIDS; Addison's disease; atherosclerosis; gout;	
KW	Graves disease; Hashimoto's thyroiditis; microbial infection; cancer;	
KW	cell proliferative disorder.	
XX		
OS	Homo sapiens.	
XX		
PH	Key	Location/Qualifiers
FT	Modified-site	138
FT	Modified-site	/note= "phosphorylation site"
FT	Modified-site	174
FT	Modified-site	/note= "phosphorylation site"
FT	Modified-site	251
FT	Modified-site	/note= "phosphorylation site"
FT	Modified-site	376
FT	Modified-site	/note= "phosphorylation site"
FT	Modified-site	574
FT	Modified-site	/note= "phosphorylation site"

XX	Key	Location/Qualifiers
EH	Misc-difference	620..640
EH	Misc-difference	/note= "transcription activation domain"
EH	Misc-difference	566..573
EH	Misc-difference	/note= "potential nuclear targeting motif"
EH	Misc-difference	573..580
EH	Misc-difference	/note= "potential nuclear targeting motif"
EH	Misc-difference	576..583
EH	Misc-difference	/note= "potential nuclear targeting motif"
EH	Region	460..764
EH	Region	/note= "COOH terminal region"
EH	Region	1..175
EH	Region	/note= "NH2 terminal with 4 Asp-linked glycosylation sites"
XX	W09420616-A.	
XX	15-SEP-1994.	
XX	04-MAR-1994;	.94WO-US01860.
XX	05-MAR-1993;	93US-0026045.
XX	(SCHW/) SCHWEINFEST C W.	
XX	(USSH) US DEPT HEALTH & HUMAN SERVICES.	
XX	Papas TS, Schweinfest CW;	
XX	WPI: 1994-303023/37.	
XX	N-PSDB: AA071399.	
XX	Isolated polypeptide that is down regulated in colon	
XX	adenocarcinomas and adenomas - is used as an indicator of tissue	
XX	abnormality.	
PS	Claim 3; Page 25-27; 41pp; English.	
CC	DRA can be used as antigens for the production of monoclonal	
CC	antibodies, which may then be used diagnostically for detecting	
CC	the presence/absence of DRA polypeptide in tissue sample. Absence	
CC	of DRA protein indicates tissue abnormality.	
XX	Sequence 764 AA:	
50	Query Match	28.8%; Score 1116; DB 15; Length 764;
	Best Local Similarity	33.9%; Pred. No. 7.9e-104;
	Matches 249; Conservative 142; Mismatches 275; Indels 68; Gaps 11;	
OY	7 RRVVRAAYSLTFPDEPEKKDRTPVG-EKLRNAFRCSSAKIKAVVGLPLVLSMLPKY 65	
DB	8 qyivrrpyvstnafeenhkkgtrhktfldnlkvccscspqakakivislfrjaspwlpay 67	
OY	66 KIKDIYIPDLGLSGSGSIQVPOGNAMAFALLANLPAVNGLYSSFPFLTYFFLGGVHOWVP 125	
DB	68 rlkenwlsdivegstigtvaivqiafalfalldvdpvgdgysafefailylffgtsrhlsv 127	
OY	126 GFNAVISTILVG----NICQLAPESKFOVF---NNATNESYVDTLAMEAEERLHVSATLAC 178	
DB	128 gpfllismvvglaevsgavskayvdrnatclgipnonsnslidd--ervvraaaasvly 184	
OY	179 LVAIIOMGLGFMQGFQFVAIYLSSEFIRGFMTAAGIOILISVLYKTFGLITPSTYGPGSIV 238	
DB	185 lsglllqiafqlgltifvnyvlysesltsqgttaaaahvlysgklifqltlvpshtcpslf 244	
OY	239 FTFIDICKNLPHPTNATSLAFALISGAFVLVELNARVNHKIRFPPIEMIVVVVATAIS 298	
DB	245 kvlvsvfsgiektnadlvtalvllvsvivneigrfdkprlprieifmtvtaags 304	
OY	299 GCKMKPKTKHMOIVGEIQRGFPFPVSPVVSQKKDMIGTAFSLAISYVINLAMGRTLANK 358	
DB	305 ygcdfenrkfvaavvudmpgfgfppltpdvctfnvqdcfgiaawaifavaifvaasyisik 364	

[illegible]

CC the above vector as the active component. The sulfate transporter
 CC gene-containing vector is high in expression efficiency. This sequence
 CC represents the rat DTOST protein used in the method of the invention.
 XX
 XX
 SQ Sequence 739 AA:

Query Match 25.2%; Score 974.5; DB 20; Length 739;
 Best Local Similarity 31.7%; Pred. No. 1.7e-89;
 Matches 219; Conservative 140; Mismatches 280; Indels 51; Gaps 8;

OY 33 VGEKLRNFRCSAKIKAVFGLPLVLSLPRKIKDYIIPDLGLSGSIQVPGMAF 92
 DB 70 VNRKIKQSCGCAKIRNIFQIFPLVLPKPYDLKKNIGDMGSLVYGILLVPQSLEY 129
 OY 93 ALLANLPAVNGLSSEFPFLTFEFLGQVHWVPGFAVSIIVGN-----CLQL-A 143
 DB 130 SLLAGQEPYGLYTFSAIFYLFGTSRHSVGLFGICMLGVEDLHKACPDLOT 189
 OY 144 PESKFOVENN--ATNESYVDTRAMEAEERLHVSATLACTTAIOMLGFGQFVAIYLS 200
 DB 190 TSSSLAMFSGVGVVNHLDIGLCKSCYAIKIGSTVLFMAGYVYAMGFFGYGFSVYLS 249
 OY 201 ESFINGFMTAGLOLLISVLKIRFGLTIPSTYSGSIVTFPDKCNLPHYNASLIRPL 260
 DB 250 DALISGIVYGAFFLLISGAKYLLGISIPRNSGVSVLTWHLIFRNLKLNLCIDILTS1 309
 OY 261 ISGALVLYKELNARVMHRIPIPEMIVVAVATASGCGMPKKYHQIYGEIQRGP 320
 DB 310 LCLLYVPTKEINLEYFKSLIPAPITELIVVAATLASHFGKINENYSLAGQIPCTGM 369
 OY 321 TPVSPVSGQKMDICTAFSLATVSYVINLWAGRTLANKHGVDVDSNOEMIALGCSNFGS 380
 DB 370 PPQAPDWSLIPNVAVDATAISLIGFALIVSLEMFAKKHGYLVKANGEMAYGIFCNLPS 429
 OY 381 PFKHIVICALSVTLAVDAGAGSKSOVASLCSLVYVMTLVIGITLYPLPKSVLCAIYV 440
 DB 430 FHCILTSAAALAKLIVKESTGCTGLSALVTSIVILLVILLIAPLFGSLQKCVLIGVLTIV 489
 OY 441 NLKNSLKQLTDPYLVWRKSKLDCCIWVVSFLSPFLDPYGAAGVAFSVLVVVOGTOR 500
 DB 490 NLRGALLKFRDIPKMTIRMTLVLFVMTSSALISTEIGLLVGVCTSMFCVLLITQMP 549
 OY 501 NGYALQAVMDNDIVNPKTYNRADIDGIRITTCSPFLFANSEIFROKIVAKTYSLOEL 560
 DB 550 KSLIIGLEESSEIFESISYKNLRSGKJGKFRFLAPLYLNKECKFSAIYKKTLPNVLY 609
 OY 561 QODEFENAPPTDPNNNOTFANGTISVYITFSPDSSPAQSEPPASAEAPGEPMDMLASVP 620
 DB 610 KAAMKKAARKLKEET-----VTFHGD-----PDEVSMQISHDP- 643
 OY 621 FVTFHTLIDMSGVSFVDLGIKALAKLSTSTGKIGVKKVFLVNIHNAVNDISHGVFED 680
 DB 644 -LELHTLVLDCAIFIDTAGLHCLKVRYDYEAGIYGLIAGCNPVSVDLSAKGEY--- 699
 OY 681 GSLECKH---VPPSHDAVLFAQANARD 705
 DB 700 ----CKKEENLLFYSLSAFAESQKE 725

RESULT 4
 ID AAY08928 standard; protein: 739 AA.
 XX
 XX AAY08928;
 AC
 XX 19-AUG-1999 (first entry)
 XX
 XX Human DTOST protein.
 DE
 XX DTOST: human, expression vector; sulphate transporter; screening;
 KW bone disease; cartilage disease; treatment; cell sulphate ion intake;
 KW drug preparation.

XX Homo sapiens.
 OS JP1146790-A.
 PN 02-JUN-1999.
 XX
 XX 18-NOV-1997; 97JP-0335157.
 XX 18-NOV-1997; 97JP-0335157.
 XX (SOMU) SUMITOMO SEIYAKU KK.
 PA WPI: 1999-378999/32.
 DR Sulfate transporter gene expression vector
 XX
 PS Claim 2; Page 10-11; 22pp; Japanese.
 XX
 CC This invention describes the construction of a novel vector for sulfate
 CC transporter expression containing a DNA sequence encoding a mammalian
 CC sulfate transporter (expression product of the DTOST gene) and
 CC containing no DNA sequence of 5' translation region of the mammalian
 CC sulfate transporter gene. The invention also describes: (A) a method for
 CC screening a human bone/cartilage disease treating agent including the
 CC steps: (1) transforming an animal cell with the above vector, (2)
 CC culturing the animal cell in the presence of a sample and (3) detecting
 CC the increase in the sulfate ion intake to the cell; (B) a drug
 CC preparation for the treatment of human bone/cartilage disease containing
 CC the above vector as the active component. The sulfate transporter
 CC gene-containing vector is high in expression efficiency. This sequence
 CC represents the human DTOST protein used in the method of the invention.
 XX
 XX
 SQ Sequence 739 AA:

Query Match 25.1%; Score 972; DB 20; Length 739;
 Best Local Similarity 31.2%; Pred. No. 3e-89;
 Matches 226; Conservative 138; Mismatches 285; Indels 76; Gaps 11;

OY 3 QRPKRYVDRAVSLTLPDEFKKDRTP--VGEKLRNFRCSAKIKAVFGLPLVLS 60
 DB 49 RPYHILLERQ-----EKEDLNKEFVKKLQKNCQSPAKAKNMILGILPVYIQ 97
 OY 61 WPKYKIKDYIIPDLGLSGSIOVPGCMAFALANLPVNGIXSPFLTLYFFLGCY 120
 DB 98 WLPKYDLKKNIGIDVMSGLIYGLIYVPGSISYLLAGEPEYGLYTSFSAIYFLLGTS 157
 OY 121 HQMVPQTEFAVISILVGNIC--LQ-----LABESKFOYFNNTNESYVDTAAM 165
 DB 158 RHISVGLIFGLCLMIGETVRELQKAGYDHAHSAPSLIGVNSGSLIHLTSRDLCKSCY 217
 OY 166 EAERLHVSATLACTTAIOMLGFGPMQFVATLYSESRFGRMTAGLOLLISVLKIRF 225
 DB 218 ---ALMVGSTVFLIAGYGVVAMGFFGYGFSVYSLDALLISGFTVGASFTLSGAKYLL 274
 OY 226 LTRPSYSGSIVTFPDKCNLPHYNASLIFALISGALVVLVLENARVMHRIPIRP 285
 DB 275 LNPRTNGVSGITLWVHFNIRHKNLCIDILTSILCLLVLPTELNEHFKSKKAPRP 334
 OY 286 TEMIVVAVATAISGCGMPKKYHMOIYGEIQRGFPVPSVVSQMK---DMIGTAFSLAI 342
 DB 335 LELVVVAATLASHGKLNHENYSLAGHPIPTG---MPKQVEMNLLPSVAVDATAISL 391
 OY 343 VSYVINLWAGRTLANKHGVDVDSNOEMIALGCSNFGSFFKIHVICALSVTLAVDAGAG 402
 DB 392 IGFALIVSLEMFAKKHGYTVKANGEMAYIGFCNLIPIFFHCFTSAAALAKLIVESGTC 451
 OY 403 KSOVASLCSLVMTLVVLGITYLPLPKSVYALIAVLAKNSLKQLTDPYLVWRKSKLD 462
 DB 452 HTGISGVVATVILLVILLIAPLFGSLQKSVGLVLTIVNLRGALKKFRDIPKMSWISLMD 511
 OY 463 CCIMWVSFLSPFLPYGAAGVAFSVLVVVOGTORNGVALQAVMDTIYVNPRTYNR 522

Db 512 twivwtvtnlssallsteigllivgvcifcviiltqdkpkslllgllveesevtesvsaaykn 571

Qy 523 AODIIGITITTYGCPLETFANSEIRQKVIKTVASLOELQODFENAPPTDPNNQOTFANCT 582

Db 572 lqtgpgikiftfvaplylyinkeckksalylqvtynpllikwkkkaakrkikkekvvrlgsl 631

Qy 583 SVSYITTESPDSSSPROASEPPASAEAPGEPSPDMLASVPEVTEFTPLILDMSGVSFVLMGI 642

Db 632 q-----demsvqlshdp-----telhtlvidcsaqlfdtagl 664

Qy 643 KALAKLSTGYKICGVKFLVNIHQVYNDISHGCVFEEDGSLBEKH-----VFPSIHDAVL 697

Db 665 htltveirrdyaiaigylalgaqncplvtrsltngey-----ckkeeenllfysvyeama 717

Qy 698 FQAQAN 702

Db 718 faevs 722

RESULT 5

AA15272

ID AA15272 standard; Protein: 506 AA.

AC AA15272;

XX 12-OCT-2001 (first entry)

DT

XX Peptide #1706 encoded by probe for measuring cervical gene expression.

DE

XX Probe: human; microarray; gene expression; cervical epithelial cell;

KM cervical cancer.

XX

OS Homo sapiens.

XX

XX W0200157278-A2.

PN

XX 09-AUG-2001.

PD

PF 30-JAN-2001; 2001WO-US00670.

XX

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR:

XX

DR WPI: 2001-488901/53.

XX

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human cervical epithelial cells -

XX

PS Claim 27; SEQ ID No 20098; 487pp; English.

XX

XX The present invention relates to human single exon nucleic acid probes

CC (SNP: see AA110068-AA128459). The present sequence is a peptide encoded

CC by one such probe. The SNPs are derived from human HeLa cells. The SEMP

CC can be used to produce a single exon microarray, which can be used for

CC measuring human gene expression in a sample derived from human cervical

CC epithelial cells. By measuring gene expression, the probes are therefore

CC useful in grading and/or staging of diseases of the cervix, notably

CC cervical cancer.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIFO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX

XX Sequence 506 AA;

Query Match	19.0%	Score 735.5	DB 22	Length 506
Best Local Similarity	31.2%	Pred. No. 1.6e-65		
Matches 164	Conservative 105	Mismatches 212	Indels 45	Gaps 6
OY	185 MCLGEMQGFVAIVYISESEFIRGFMPRAGQILISYKTYIEGLTIPSYTSPGSIPTFIDI	244		
Db	1 vsmgffiqvgyvsvysdallsgfvvgastllttsqakylglnlprtgvgalltwtlhv	60		
OY	245 CKNLPRTNIASLIFALISGATLVYKELNARYMHKIRPIPTREMIYVVAATISGCKMP	304		
Db	61 fznlnhtncltdltsllcllvpilpckelnehfkskikapietlvvvaatlshgkin	120		
OY	305 KKYHHQIYVEIQGFPTPSVYSQK---DMIGTAFSLAIVSYVNLAMGRTLANKHGX	361		
Db	121 enynsliaghlptgt---mpkpvpevnllipsavaddatalslfgalvalisemfakkhgy	177		
OY	362 DVDSDNEMALCCSNFEGSFPIHNYICALSVTLAVDAGGKSOVASLCSLVMTIMLV	421		
Db	178 tvkangemyaigfcmilipisfthcfttsaalaktlvkstgchltqisgvvtalvlllvllv	237		
OY	422 LGIILYPLPKSVGLGALIANVKNLSKQLTDPYILMRKSKLDCCIVVSVFLSSFFLSLPYG	481		
Db	238 lapiflsyiqksylyvltvlnlrgalrktfrdpkkmwsismtdvliwfvthlssallsteig	297		
OY	482 VAVGAVSVSLVAVVFOGFNGVYALQVMTDLYVNPKNTPNRADLOGIKIITRYCSPLYFA	541		
Db	298 llvygcfsifvcyillrtqkpsllglveesevfesyayknlgkpykllfrtvaelyyi	357		
OY	542 NSEIFRQKIYAKTVSLQELQODFENAPPTDPNNNQTFRANGTSVSYITFSPDSSSPAQSEP	601		
Db	358 nkccfsalsykgvtvpnlilikvawkkaarkkikekvllgigq-----demsyqlnd	409		
OY	602 PKSAEKGSPSDMLASVPFVYTFHTLILDMSGVSFVLDMLGIALAKLSSTYKIGIKVFL	661		
Db	410 p-----lehlvtidcsaigfidtagihltikvevrrdyeaigivll	450		
OY	662 VINIAQVYNDISHGGEFEDGSLECKH----VFPSIHDLVLPQAQN	702		
Db	451 aqcnprvtvrdslngcy-----ckkeenllfysyemaataevs	489		
RESULT	6			
AAM27737				
ID	AAM27737 standard; Protein; 506 AA.			
XX	AAM27737;			
AC				
DT	17-OCT-2001 (first entry)			
XX	Peptide #1774 encoded by probe for measuring placental gene expression.			
DE	Probe: microarray; human; placenta; antenatal diagnosis;			
XX	genetic disorder.			
KM				
OS	Homo sapiens.			
XX				
PN	WO200157272-A2.			
XX				
PD	09-AUG-2001.			
XX				
PE	30-JAN-2001; 2001WO-US00663.			
XX				
PR	04-FEB-2000; 2000US-0180312.			
PR	26-MAY-2000; 2000US-0207456.			
PR	30-JUN-2000; 2000US-0608408.			
PR	03-AUG-2000; 2000US-0633266.			
PR	21-SEP-2000; 2000US-0234687.			
PR	27-SEP-2000; 2000US-0236359.			
PR	04-OCT-2000; 2000GB-0024263.			
XX				
PA	(MOLE-) MOLECULAR DYNAMICS INC.			

XX	Homo sapiens.
OS	
XX	
XX	
PN	W0200157270-A2.
XX	
XX	
PD	09-AUG-2001.
XX	
XX	
PF	29-JAN-2001; 2001WO-US00661.
XX	
PR	04-FEB-2000; 2000US-0180312.
PR	26-MAY-2000; 2000US-0207456.
PR	30-JUN-2000; 2000US-0608408.
PR	03-AUG-2000; 2000US-0632366.
PR	21-SEP-2000; 2000US-0234687.
PR	27-SEP-2000; 2000US-0236359.
PR	04-OCT-2000; 2000GB-0024263.
XX	
XX	
PA	(MOLE-) MOLECULAR DYNAMICS INC.

FI	Fein SS, Nathan DN, Clich W, Nalin DN,
XX	
DR	WPI, 2001-476286/51.
XX	

PS Claim 27; SEQ ID No 11757; 322pp; English.

CC The present invention relates to novel single exon nucleic acid probes
CC (see AAI00010-AA110067). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for measuring human gene expression in

CC nucleic acid expressed in the human breast. The probes are useful for
CC predicting, diagnosing, grading, staging, monitoring and prognosing
CC diseases of the human breast, particularly those diseases with polygenic
CC aetiology. The diseases include: breast cancer, disorders of development

CC inflammatory diseases of the breast, monocyclic changes, proliferative
CC breast disease and non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO

XX	Sequence	506 AA;
SQ		

Query Match	19.08;	Score 735.5;	DB 22;	Length 506;
Best Local Similarity	31.28;	Pred. No. 1.6e-65;		
Matches 164;	Conservative 105;	Mismatches 212;	Indels 45;	Gaps 6

QY 185 MGAGMGCGFVAATLYLSESTIRGFMTAAGILAILSLVKTYIGTLTTSYTSGSIVTFEIDI 244
:
Db 1 vvangffqygfsvlylsdallsgfvtgasftllcsqakyllglnlprtnygsliltcwlhv 60

```
QY      245 CKNLPHTIASLIFALISAFELVEKLELARMYMHIIRPTETEMVWVAATASGCKMP   304
           ::::: :::::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db      61 frnihktrlcdiltsllcrlvllptkelnehfksklkapipetelvvvaatlaasftgkln 120
```

305 KAHIMQUGELQKKEPRTPSPVNSQMK---DMIGTAFSLAIVSYINLAQRORILNKKHGX 361
 :|:| | | | | :|::|:|::|
 Db 121 enynsslaqhlprgf---mpkpvrwmlipsvavdalalslglfalfvtslsemfakkhgy 177

Db 178 tvkangemya1gfcni1psffhncfttsaalaktlvkescfgtqsgvavtalvllvllv 237

238 iaplfyslqksvlgvltivnlrgalrkfrldpkmsisrmdtviwfvmlssallsteig 297

```

: ||| ||: |: ||| | | :||| : |||:
298 llvgvcfsifcvllrtqkpkssllgllvesevfesvsayknlgctkpgiklfrfvaplyyl 357

```

Oy	542	NSLFFROKVIKATYSLQELQODFENAPPTDPNNQGTANGTSVSYITFFSPDSSPAPQSEP	601
Db	358	nkccskssalykqlyvnpillikvawkkakrkikxkvvl1991q-----demsvglshd	409
Oy	602	PASAEAPCEPSPDMLASVPPPEFTFTLLIDMSGVSFVDMGLKALAKLSSTYGRKGVVFL	661
Db	410	p-----lehtltyidcsaigfltagihlckeyrrdyeaig1qvl1	450
Oy	662	VNIHAQVYNDISHGCVFEDGSLSECKH-----VPPSIHDAVLEFAQAN	702
Db	451	agcnpvtvrdsitlungey-----ckkeeenllfysyveamafeavs	489
RESULT	8		
ID	AAM42394		
AA	AAM42394	standard; Protein: 143	AA.
XX	AAM42394;		
XX	22-OCT-2001	(first entry)	
DE	Human	polypeptide SEQ ID NO 127.	
KW	Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;		
KW	Immunosuppressive; antiinflammatory; anti-HIV; antibacterial; antiviral;		
KW	antiparkinsonian; antischizoid; antianemic; antiallergic; cancer;		
KW	antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;		
KW	antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;		
KW	antiparasitic; cardiac; immune disorder; cardiovascular disorder;		
KW	neurological disease; infection; nephrotropic; gene therapy; vaccine.		
XX			
OS	Homo sapiens.		
PN	MO200155449-A1.		
XX			
PD	02-AUG-2001.		
XX			
PE	17-JAN-2001; 2001MO-USO1346.		
XX			
PR	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180628.		
PR	19-MAY-2000; 2000US-0205315.		
PR	07-JUL-2000; 2000US-0216880.		
PR	14-JUL-2000; 2000US-0218290.		
PR	14-AUG-2000; 2000US-0225447.		
PR	01-SEP-2000; 2000US-0229343.		
PR	06-SEP-2000; 2000US-0230437.		
PR	08-SEP-2000; 2000US-0231243.		
PR	25-SEP-2000; 2000US-0234997.		
PR	29-SEP-2000; 2000US-0236367.		
PR	13-OCT-2000; 2000US-0239937.		
PR	08-NOV-2000; 2000US-0246476.		
PR	08-NOV-2000; 2000US-0246477.		
PR	08-NOV-2000; 2000US-0246525.		
PR	08-NOV-2000; 2000US-0246526.		
PR	08-NOV-2000; 2000US-0246528.		
PR	17-NOV-2000; 2000US-0249210.		
PR	17-NOV-2000; 2000US-0249211.		
PR	17-NOV-2000; 2000US-0249214.		
PR	17-NOV-2000; 2000US-0249265.		
PR	01-DEC-2000; 2000US-0250160.		
PR	01-DEC-2000; 2000US-0250391.		
PR	05-DEC-2000; 2000US-0251030.		
PR	05-DEC-2000; 2000US-0251988.		
PR	05-DEC-2000; 2000US-0256719.		
PR	06-DEC-2000; 2000US-0251479.		
PR	08-DEC-2000; 2000US-0251989.		
PR	08-DEC-2000; 2000US-0251990.		
XX	11-DEC-2000; 2000US-0254097.		
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
PI	Rosen CA, Barash SC, Ruben SM;		

XX WPI: 2001-4/6225/51.
DR N-PSDB: AA162799.
XX
PT Novel plasma membrane associated proteins useful for diagnosing,
PT treating, preventing and/or prognosing disorders related to the
PS proteins, including cancer, immune response and neuronal disorders -
XX
PS Claim 11: SEQ ID NO 127: 532pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (AA162752-AA162961) and proteins
CC (AA162347-AA162415) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (anti)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 143 AA:

Query Match	15.5%	Score 601.5;	DB 22;	Length 143;
Best Local Similarity	91.0%	Pred. No. 8.9e-53;		
Matches 122;	Conservative	3;	Mismatches 8;	Indels 1;
				Gaps 1;
Qy	107	SEFPILTFEFLGCVHQMVGPTPAVISILVGNICLOLAPESKPFQVENNATNESYVDTAAME	166	
Db	9	sfllpiltffixgsvhgmvygtfavlsilvgnclclapeskqvfnnatnesyvttaame	68	
Qy	167	AERLHVSAITLACTTAIIQWGLGFMQGFVAIYLSSEFIRGEMTAAGLOILISVLKYIFGL	226	
Db	69	aerlhvsatlactlxtklqmgjlgfmqfgfvaalylsefifrgfntaaglqlilsvlkyifga	128	
Qy	227	-TIPSYTGSGSTVF	239	
Db	129	gqepchswprslaf	142	
RESULT 9				
AAV44945				
ID	AAV44945	standard; Protein; 593	AA.	
XX	AAV44945;			
AC				
XX				
DT	23-MAY-2000	(first entry)		
XX				
DE	Wheat sulphate permease-2.			
XX				
KW	Sulphate Permease; sulphate assimilation protein; wheat; probe;			
KW	mapping; marker; plant breeding; chimeric gene; transgenic plant;			
KW	antibody; screen.			
XX				
OS	Triticum aestivum.			
XX				
PN	WO200004154-A2.			
XX				
PD	27-JAN-2000.			
XX				
PF	13-JUL-1999;	99WO-US15810.		
XX				
PR	14-JUL-1998;	98US-0092833.		
XX				

PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Allen SM, Falco SC, Thorpe CJ;
 XX
 XX WPI: 2000-195025/17.
 DR N-PSDB: AAZ44945.
 XX
 XX Nucleic acid fragments encoding sulfate assimilation proteins in plants
 PT and seeds useful as probes for isolating cDNAs and genes encoding
 PT homologous proteins, in producing transgenic plants -
 XX
 PS Claim 6; Page 59-60; 79pp; English.
 XX
 XX The present amino acid sequence is the wheat sulphate permease, a
 CC sulphate assimilation protein. This sequence is obtained from wlm4
 CC library, clone wlm4.pk0016.all, derived from wheat seedlings, 4 hours
 CC after inoculation with *E. graninis*. It has 73% sequence identity to
 CC Arabidopsis thaliana sulphate permease (gi 2626753).
 CC This sequence is used as a probe to isolate other plant sulphate
 CC assimilation proteins, for genetic and physical mapping of related genes
 CC and as markers of traits linked to the gene. This is useful for plant
 CC breeding and to construct chimeric genes, used to create transgenic
 CC plants with altered levels of sulphate permease. The sulphate permease
 CC peptides are useful for producing antibodies, that are used to screen
 CC and isolate cDNA clones.
 XX
 XX Sequence 593 AA:
 Query Match 14.9%; Score 576; DB 21; Length 593;
 Best Local Similarity 24.2%; Pred. No. 3.2e-49;
 Matches 161; Conservative 127; Mismatches 248; Indels 128; Gaps 13;
 QY 56 LPVLSMLPKRYKIDYIPDLGLSGSIOYPOGMAFALLANLPAVNGLYSFFPLTYF 115
 Db 21 lpcldammsytwkedfegfdlaagltvgymalpamsyakiagldhpylytfgfvpalfya 80
 QY 116 FLGGVHQWVPGTFAVISILVGNICQLAPESKQYFNNAATNESYVDTAMEAERLVSAT 175
 Db 81 lfgsarqlavgpvalvsllvsnvlgiv-----nsseelyelall----- 121
 QY 176 LACLTATIQMGIFMGQFQFVAIYSESFIRGPTAAGQILISVAKYIFGLT----- 227
 Db 122 lafmvglliecimalrltqwlrltishvslgftsaalvlgldkylfgyvstsrskll 181
 QY 228 -----IPSYTGP-----GSIYFTFDICKNLPHTNASTIFALISGAFVLVKEDNA 274
 Db 182 pliesitgldqfswpftvmsafivlllmkklgktn-klrlftrasgpltav----- 235
 QY 275 RYMKIRPPIPTEMIVVAVATAISGCKMPKRYHMOIVGEIQRGPT- PVSFVVSQMKDM 333
 Db 236 -----lgtlfvklfrfais-----vvgeipqgfpstisipgfehlnsl 274
 QY 334 IGTAESLAIYGVINLNGRILANKHGYDVDSNOMIALGCSNFGSFFKIHVICALSV 393
 Db 275 mptalltgvalllesvgalakaaingyelsnkelfgljsnlogsfisaypalsfsr 334
 QY 394 TLAVDAGAGSGOVASLGVSLVWIMLVGLIYLPKRSVICALAVNMLKNSLKQULTPY 453
 Db 335 savnhesgaktglsinglllscallfntplfdipgcalaatv-isaavglvdyeeat 393
 QY 454 YLMRKSRLDCICIVVVSFLSFFLSLPYGVAVGAFSLVAVFQTFGRNGYALAQVMDTI 513
 Db 394 flwgdkkdfellwmtfcltllfgtelsglvlgvgsfalsfhlesanplavlgtripttv 453
 QY 514 YVNRPTYRADDIQIKITTCSPLYFANSEIFRKQVIAKTYVLSLOELOODENAPPTDPN 573
 Db 454 yrnltqypeaytynglivvrvdapyfanlsyikrlreyelkl-----pn 499
 QY 574 NNQTPANTSVSYITFSPDSSSPAQSEPPASAEARGEPSDMLASVPPVTFPHLLDMSG 633
 Db 500 snrtpdyg-----ryfvllemp 518

QY 634 VSFVDLMGKIALAKUSITYGKIGVAVFLVNIHAQVYNDISHCGRF-DQSECKHVPFSI 692
 Db 519 vlyidssavgalckdhgeykardiqalamprryghlllsregldimdgawc---fvr 575
 QY 693 HDVAF 696
 Db 576 hnav 579
 RESULT 10
 ID AAY44943
 ID AAY44943 standard; Protein: 680 AA.
 XX
 XX AAY44943;
 XX
 XX 23-MAY-2000 (first entry)
 XX
 DE Soybean sulphate permease-2.
 XX
 KW Sulphate permease; sulphate assimilation protein; soybean; probe;
 KW mapping; marker; plant breeding; chimeric gene; transgenic plant;
 KW antibody; screen.
 XX
 OS Glycine max.
 XX
 PN WO200004154-A2.
 XX
 PD 27-JAN-2000.
 XX
 PF 13-JUL-1999; 99WO-US15810.
 XX
 PR 14-JUL-1998; 98US-0092833.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Allen SM, Falco SC, Thorpe CJ;
 XX
 DR WPI: 2000-195025/17.
 XX
 DR N-PSDB: AAZ50489.
 XX
 PT Nucleic acid fragments encoding sulfate assimilation proteins in plants
 PT and seeds useful as probes for isolating cDNAs and genes encoding
 PT homologous proteins, in producing transgenic plants -
 PS
 PS Claim 6; Page 53-55; 79pp; English.
 CC
 CC The present amino acid sequence is the soybean sulphate permease,
 CC a sulphate assimilation protein. This sequence is obtained from sf11
 CC library, clone sf11.pk0043.g10, derived from soybean immature flower.
 CC It has 75% sequence identity to Arabidopsis thaliana sulphate permease
 CC (gi 2265885). This sequence is used as a probe to isolate other plant
 CC sulphate assimilation proteins, for genetic and physical mapping of
 CC related genes and as markers of traits linked to the gene. This is
 CC useful for plant breeding and to construct chimeric genes, used to create
 CC transgenic plants with altered levels of sulphate permease. The sulphate
 CC permease peptides are useful for producing antibodies, that are used to
 CC screen and isolate cDNA clones.
 XX
 XX Sequence 680 AA:
 Query Match 14.6%; Score 566; DB 21; Length 680;
 Best Local Similarity 24.9%; Pred. No. 4.1e-48;
 Matches 183; Conservative 128; Mismatches 293; Indels 132; Gaps 19;
 QY 4 PREPRVDRAYSL-TLFDDEFKKDRITYPYGKLRNARFSSAKIRAVVFGL---LPV 58
 Db 47 pppqgffkslysketlfpddpirltqfknkpsakf-----mlgqfffpil 92
 QY 59 LSWLPKRYKIDYIPDLGLSGSIOYPOGMAFALLANLPAVNGLYSFFPLTYFPLG 118
 Db 93 fewapkyctfg-fikadliaglitiaslaipqgislyaklanlpllglysfippliyamng 151

```

OY 119 GVHOMVPGTFAVISILVGNICLOLAPESKFOVENNATNESYDTAAEAEPLHVSATLAC 178
    : 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 152 sardlavgtavagslmgs-----mlsna-----vdpneqpkjlylhafatcl 194
OY 179 LFAIHWGLGFMOGFVAIYLSSEFIRGFMTAGLOLILSVLKIFGLTIPSYSGPSIV 238
    : 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 195 fagvfgaalylfifglivdflshatltgmgaaavciglqslslyl--ehthgadi 252
OY 239 FREDICKNLPHNTIASLIFALISGAFLVVKELNARYMKRREPITEMIVVVA--TA 296
    : 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 253 svmsvftqchewrweaavlgcvffl-----lstryfakkr---prffvwsamapls 304
OY 297 ISSGCKM-----PKRYMIOVGEIQRGFPFP-----VSPVVSOMKDMIGTAFSAIVS 344
    : 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 305 vllgsllyvftchaekbhvgeigklglnpsltnlyfvsapymt---lavktgtvgvlls 361
OY 345 YVINLWAGFLANKHGYDVDSNOEMIALGCSNFGSPFKIHVICALSVTLAVDAGAGKS 404
    : 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 362 laeglavgrfamyknyndgnkemaiglmvvgstscyltqpftrsavmynaagckl 421
OY 405 QVASLGVSLVMTMLVIGIYLPKPSVGLALIAVNLKNSLKOLTPPYILMRKSKLDCC 464
    : 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 422 aasnllmslavmlltllfltpfhytprlvlsall--vsamigldysaahlfkvdckdfv 480
OY 465 IWWVSFLSSFFLSPYGVAVGAVSVLVVVFQTOFRNGYALAOVMDIYVNPRTYNAQ 524
    : 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 481 vcmssaylgvvgfagvelglvalvslvrlvllflarprtfvlgmnsvlyrnehvgnak 540
OY 525 DIQIKITTCSPLYFANSEIFROKVIKTVSLOELQODEFNAPPTPNNTANG--TS 583
    : 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 541 hvgmllleldaplyfianasylericrw-----ldeeerlkatgets 584
OY 584 VSYITFSPDSSSPAQSEPPASAEAPGEPSSDMLASVPFVTFHTLLDMSGVSFVDMGIR 643
    : 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 585 lgyv-----ldmsavgnidtsais 604
OY 644 ALAKLSITGKIGKVLVNIHAQVYNDISHGVFEDGSELECKHVFPSSIHDAVFAQANA 703
    : 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 605 mleeavkklterrelqlvlpvsevmklnkskf--qnhlgkwllyltveeavagcnfnl 662
OY 704 R-----DVTFGHN 711
    : 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 663 raaktnpkdtegevn 678

```

RESULT 11

AA44935
ID AAY44935 standard; Protein; 688 AA.

```

XX AC AAY44935;
XX DE 23-MAY-2000 (first entry)
XX DE Corn sulphate permease-1.
XX DE Sulphate Permease; sulphate assimilation protein; corn; contig; probe;
KM mapping; marker; plant breeding; chimeric gene; transgenic plant;
XX KM antibody; screen.
XX OS Zea mays.
XX PN WO200004154-A2.
XX PD 27-JAN-2000.
XX PF 13-JUL-1999; 99MO-US15810.
XX PR 14-JUL-1998; 98US-0092833.
XX PA (DUPO ) DU PONT DE NEMOURS & CO E. I.
XX PI Allen SM, Falco SC, Thorpe CJ;

```

```

DR WPI; 2000-195025/17.
XX N-PSDB; AAZ50481.
XX PT Nucleic acid fragments encoding sulfate assimilation proteins in plants
PT and seeds useful as probes for isolating cDNAs and genes encoding
PT homologous proteins, in producing transgenic plants -
XX Claim 6; Page 39-41; 79pp; English.
XX The present amino acid sequence is the corn sulphate permease, a
CC sulphate assimilation protein. This sequence is from a contig composed
CC of clones cbn10.pk0062.b10, ccoln.pk0081.h21, ccoln.pk0092.12,
CC cscic.pk0005.j3, p0004.cb1e58r, p0089.cscic19r, p0094.cssg12r,
CC p0121.cifmx30r and p0128.cpicz09r from cbn10, ccoln, cscic, p0004,
CC p0089, p0094, p0121 and p0128 libraries, respectively. It has 80.2%
CC sequence identity to Sporobolus stapifolius sulphate permease (q11907270).
CC This sequence is used as a probe to isolate other plant sulphate
CC assimilation proteins, for genetic and physical mapping of related genes
CC and as markers of traits linked to the gene. This is useful for plant
CC breeding and to construct chimeric genes, used to create transgenic
CC plants with altered levels of sulphate permease. The sulphate permease
CC peptides are useful for producing antibodies, that are used to screen
CC and isolate cDNA clones.
SQ Sequence 688 AA:
Query Match 14.6%; Score 566; DB.21; Length 688;
Best Local Similarity 25.0%; Pred. No. 4.2e-48;
Matches 173; Conservative 123; Mismatches 231; Indels 146; Gaps 18;
OY 18 TLFDEFEKKDRTPVGEKLNARFCSSAKIKAVFGLLPVLSMIPKRYIKDYIIPDLG 77
    : 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 85 tffpddpfrakgqppqtqvmavry-----lflldwvpsyl-slfksdva 132
OY 78 GLSGSIOVPGGMARFALLANPAVNGLYSSFFPLITFELGCVHOMVPGTFAVISILVGN 137
    : 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 133 gltlaalaipqisayaklasiprliglystfvpmyavvgssrdlavpvaslslmgs 192
OY 138 ICID-LAPESKFOVENNATNESYDTAAEAEPLHVSATLCLTAIIOMGLGFMQGFVA 196
    : 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 193 mlrquavptaeptlf-----lqlatstclfaglyqasagllrlglfvl 234
OY 197 IYLSSEFIRGFMTAGLOILISVLKYIFGLTIPSYTGSGSIVTFIDICKNLPHNTIASL 256
    : 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 235 dflskatlvgfmagaalivalqklgllg--lvnftemgypmasv---fhltsews 289
OY 257 IFALISGAFLVVKELNARYMKRIRP-----IPREMIVVVAIAAISGCG 301
    : 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 290 qlilmgvclfvfl--lsarhv-slrwpklfwsacaplasvlscllvflfka----- 339
OY 302 KMPKKYHMOIYGEIQRGFPFPVSPVVSOMKDM-----IGTAFSLAIVSYINTL---AM 351
    : 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 340 ---qnhglsitglqkgnlp-----swdkllldtlaylglmtkgtlytglstleglav 390
OY 352 GRTLANKHGYDVDSNOEMIALGCSNFGSPFKIHVICALSVTLAVDAGAGKSOVASICV 411
    : 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 391 gftfaslkdygidgnkemmaiglmvvgstscvyltgafarsavnhnagacklamsnvm 450
OY 412 SLVVMITMLVIGIYLP-----PKSVGALLAVNLKNSLKOLTPPYILMRKSKLDCCIWV 467
    : 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 451 altvmvltl---flmpfvytprnvlgall--laavglldfpavynhlwkmkndfvlvcv 505
OY 468 VSFLLSFFLSPYGVAVGAVSVLVVVFQTOFRNGYALAOVMDIYVNPRTYNAQIDIO 527
    : 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 506 catagvlfisvgeglajalavgisifrvlmgltcrpkmmvqgnikgldlydhlyeaqgvs 565
OY 528 GIKITTCSPLYFANSEIFROKVIKTVSLOELQODEFNAPPTPNNTANGTSVSYI 587
    : 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 566 gflilaaleaplnfansynlner-lkrvleesfedg----- 600
OY 588 TFSFDSSSPAQSEPPASAEAPGEPSSDMLASVPFVTFHTLLDMSGVSFVDMGIRKALAK 647
    : 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

```



```

PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-01494175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151086.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156548.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161982.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 14.28; Score 549; DB 21; Length 631;

Best Local Similarity 26.0%; Pred. No. 1.9e-46;

Matches 17; Conservative 120; Mismatches 231; Indels 126; Gaps 20;

```

QY 18 TLDFDEEKDRTPVGEKLNARFCSAKIKAVFGLPVLWLPKIKIRDIIPDLG 77
DB 25 ttfddpdrgrtgrgpnrtkl-----lraagy-lfpllwpeyfsf-sllksdvvs 72
QY 78 GLSGSIOVPOGMAFALLANI.PAVNGLYSSFFPLLTFFLCGVHOMVPGTFAVISILVGN 137

```

```

DB 73 gltiaslaipglisyaklaanlppivglystfpllvavilssrdlavpvaslailgs 132
QY 138 ICIOLAPESKFOVFNNATNESYVDPAAMEAERLAHSAFLACTTAIIONGLFMOPGFVAI 197
DB 133 mlrq-----qv-----spvddpvflq-latsstf--faglfqslgllrlgflid 175
QY 198 YLSESEFIRGEMTAGLOILISVLKYIFGLITPSTYNGPSIVFTFDICKNLPHTN----- 252
DB 176 flskatlifmggaalivslgqlkgllgt--hftkmsvvpvlssv---lqhtneweq 230
QY 253 --IASLIFALLISGARVLVVKELNARYMKIRPPIETEMIVVVAIAISGGCKMPKRYIMO 310
DB 231 flvmgvctll-----flistrhsmkkpklfwsagapilsvlclvfvtr-aerhgis 285
QY 311 IVGEIORGFPTPVSPVVSQMKDM-----IGTAFSLAIVSYVINLMMGRTLANHG 360
DB 286 vlgkipeglpnp-----swmlgfhgshlavlaktglvtglvstleglavgrtaakn 339
QY 361 YDVSNOEMIALGCSNPFSGFEKIHVICALSVTLAVDAGAGKSOVASLVCVLYVMITML 420
DB 340 yhvdkmekmialglmrvvgsalscytlgafersavvnaagaktavslvmsvcmvcll 399
QY 421 VLGIYLYPL---PKSVIGALLAVNLKNSIKQLTDPYILMRKSKIDCCIMVVSFLSFFL 476
DB 400 ---flmpifeytpnvigall-vcavigllidipachlwkdktfdlvmlocaifgvifl 454
QY 477 SLPYGAAGVAFSVLVVFOQFRNGYVALACVMDIDIVNPFYRARAADIOGIRITGCS 536
DB 455 svnglalaivglstfkilmqvtrpkmvlmnglpgcdlrydlhyheagrlpgflvlsles 514
QY 537 PLYFANSEIFRQKVIKTVLSLOELOQDFENAPPTDPNNNGTPANGTSVSYTFSPPSSSP 596
DB 515 pvnfansnyltert---srwleecnee-----qfillemsavgydvtngvsffkelkktakkd 579
QY 597 AQSEPPASAEAPGERSDMLASVPVTFHTLILDMGSVSVVDLMKIKALRSTYCKIG 656
DB 599 -----eeqekhsj-----qfillemsavgydvtngvsffkelkktakkd 579
QY 657 VKVFLVNIHAQV 668
DB 580 ieltvfmplsey 591

```

RESULT 13
AAG31621
ID AAG31621 standard; Protein: 658 AA.

AC AAG31621;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SRQ ID NO: 38004.

KW Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140685.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999;

PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144337.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144633.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148319.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0158029.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.

PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 14.0%; Score 543.5; DB 21; Length 658;
Best Local Similarity 24.9%; Pred. No. 7.5e-46;
Matches 180; Conservative 127; Mismatches 274; Indels 141; Gaps 23;

OY 4 PRRPVNRAAYSL--TLF-DDEFKKDRTYVGEKRNAPFCSSAKIKAVFGL--LP 57
DB 25 pppqpfklsqysvketlppddpfrq-----fkngnstr-----kfvlglyklyflp 69
OY 58 VLSWLPKTKIKYIIPDLGLSGSIOVPGCMARFALLANLPAVNGLYSFPPLTYFFL 117
DB 70 lfwaprynlk-fkfsdlagltiaslaipglsyaklanlppilglsyafpplvyavl 128
OY 118 GGVHGVNPTFVVISILVGNICQLAPESKQVFNFNATNESYVDFAAMEAR----LIV 172
DB 129 gsrdlavgltvvaalllg---amlske-----vdaekdpklylhl 166
OY 173 SATLACLAAIOMGLGFMOGFVAIYLSSEFTRGFTAGLQILSLVKYIRGLTPSYT 232
DB 167 atfatffagvlasigiflglfvdfshatlvgfmggaatlsvlqklgltglt-khtf 224
OY 233 GPGSIVFTFIDCKNIPHNINSLIFALISGAFLVYLKELNARYMKIRPPTMTIVVY 292
DB 225 dstdvlsvmrsvfsqtlhewres---gvlgcgflfll--lstry-----fsikpkkffwv 274
OY 293 VATTA-----ISGCGKM---PKYHMOIYGEIORGFPPTV-----SPVVSQWMDMG 335
DB 275 aamapltsvlsgllvyfthaeihyqvigdlkgl-npissdliiftspms---tavk 330
OY 336 TAFSLAIVSYVINLARGTLANKHGYVDVSNQEMIALGCSNFGSFPFKIHVTCALSVL 395
DB 331 tgljltgilaegaavgfsfamfkynidgnkemiafgmynvgstfcylyltgptsrsa 390
OY 396 AVDGAGGKSQVSLCVSLVVMITMLVGLIYLPPLKSVUGALIAVLNKLKSLQOLDPYVL 455
DB 391 vynyagckltamenlymalavmtllflfplfhytplvlsail--lsamjgldygaathl 449
OY 456 WRKSKLDCCIMVVSFLSPFLSPYGAVAVFSVLVWVFOQFRNGVALAOVMDTDIV 515
DB 450 wvvdkfdfivcmayygvvfgsvgeigivavaalstarlllffsrpktavkgnpmismiy 509
OY 516 NKTYNRAODIGIKITICYCSPLYFANSEIFRQKVIKTAKTYSLOELOQDENAPPTDPNN 575
DB 510 nteqpsstctvgjllleldapilyfanaasylerilr-----ideeee 553
OY 576 QTPRANG-TSVSYITTSPPDSSPFAQSEPPASAEKRGEPESOMLASVPPFVTFHLLIDMSGV 634

DB 554 rvkqsgesslgyl-----lllmsav 573
OY 635 SEVDLMGIKALAKISTYKIGKVFVLNITHAVYNDISHGVEFGDSLECKHVPSPID 694
DB 574 gnldtsqismwveikkvldirralktivlnpdkgevvvkltrskfigd-hlgkewmfltyge 632
OY 695 AV 696
DB 633 av 634
RESULT 14
AAG32200
ID AAG32200 standard; Protein: 658 AA.
XX
AC AAG32200;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 38800.
XX
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145228.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160761.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 14.0%; Score 543.5; DB 21; Length 658;
Best Local Similarity 24.9%; Pred. No. 7.5e-46;
Matches 180; Conservative 127; Mismatches 274; Indels 141; Gaps 23;


```

OY 456 WRKSKLDCCIMVVSFLSPFLSPYGVANGVAFSVLVVVFOTQFRNGYALAOVMDTIDIV 515
DB 495 WkvdKfdlvcmsayvgvfgsvveiglvaivalarillfvsrpktavkgnlpnsmlyr 554
OY 516 NKRTYNRAODIOGKIITFYCSPLYFANSEIFROKVIKTVSLQELQDFENAPPTDPNNN 575
DB 555 nteqypsrtvpglllleidapiyafanasylterllrw-----ideeee 598
OY 576 QTPANG-TSVSYITFSPDSSSPAQSEPPASAEAPGEPPSDMLASVPPFVTFHTLLIDMSGV 634
DB 599 rvkgsgeslgyl-----llmsav 618
OY 635 SFVDLMGIRAKALASSTYCKIGVAVFLVNIHAQVYNDISHGCVFEDGSLCKHVPPSIHD 694
DB 619 gnldtsgismmveikkvldiralkivlnpbgvkvklttrskfigd-hlgkewmfltyge 677
OY 695 AV 696
DB 678 av 679

```

Search completed: April 26, 2002, 09:05:16
 Job time: 207 sec

